ESTABLISHING THE RELATIONSHIP OF SYMPTOM VALIDITY TESTS WITH MEASURES OF AUDITORY AND VISUAL MEMORY

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

EVAN A. KOEHN

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2018

ESTABLISHING THE RELATIONSHIP OF SYMPTOM VALIDITY TESTS WITH MEASURES OF AUDITORY AND VISUAL MEMORY

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

BY

EVAN A. KOEHN

DISSERTATION ADVISOR: DR. ANDREW S. DAVIS

APPROVED BY:

_____	_____
Andrew S. Davis, Ph.D. Committee Chairperson	Date
_____	_____
W. Holmes Finch, Ph.D. Committee Member	Date
_____	_____
Amy J. Harden, Ph.D. Committee Member	Date
_____	_____
Eric E. Pierson, Ph.D. Committee Member	Date

BALL STATE UNIVERSITY

MUNCIE, INDIANA

MAY 2018

ABSTRACT

DISSERTATION: Establishing the Relationship of Symptom Validity Tests with Measures of Auditory and Visual Memory

STUDENT: Evan A. Koehn

DEGREE: Doctor of Philosophy

COLLEGE: Teachers College

DATE: May 2018

PAGES: 159

This study explored the relationships between common measures of symptom validity and a well-validated measure of auditory and visual memory. An understanding of the relationships could result in greater accuracy of psychological assessments or even reduced administration time if redundancy is found. Symptom validity and memory test measures were examined through descriptive statistics, multivariate regressions, and correlations. The research design and obtained data suggest that performance on memory measures cannot be predicted based upon symptom validity test performance for a sample of undergraduate students instructed to provide full effort. Furthermore, this study was unable to detect differences in the relationships between specific symptom validity measures and analogous measures of memory.

Extended Abstract

Symptom validity measurement is an important topic in psychological and neuropsychological assessment as providing accurate results is an ethical obligation (American Psychological Association, 2010, para. 1) and can significantly influence the likelihood of an examinee receiving life-changing benefits or consequences (Slick, Tan, Sherman, & Strauss, 2011). The primary purpose of the present study was to explore the relationship between two popular symptom validity measures, the TOMM and WMT, with a widely used measure of memory, the WMS-IV. This study was conducted as a better understanding of the relationship between these measures could result in increased confidence in obtained assessment results or even allow for eliminating unneeded measures from an assessment battery. A battery of neuropsychological tests, including the TOMM, WMT, and WMS-IV, was administered to 46 undergraduate students. Participants were instructed to provide full effort. Obtained data was analyzed through descriptive statistics as well as multivariate regression and correlational analyses. Nonsignificant regression findings between memory and symptom validity tests were found. Uncorrected correlations between measures were moderate to large. The findings underscore the need to carefully design studies and apply suitable statistical tests. Limited variability on symptom validity test performance, small sample size, and usage of methods to reduce false positives are discussed. The data suggests that, for this sample, symptom validity tests were ineffective in predicting memory performance. Furthermore, symptom validity tests and analogous components of memory measures do not have a differing relationship compared to other measured memory domains.

Acknowledgments

My journey to complete this dissertation and PhD began at Ball State University during the second half of 2008 although in some shape or form it began years or even millennia before then. I spent much of the last decade speaking to friends, family, and faculty about my doubt that I would even complete this journey. The attrition rates for PhD programs are staggering which is not surprising given the sheer amount of time that passes and the difficulty of balancing school with the demands of life. If you are reading this right now, I am among one of the extremely lucky people to have made it. Chances are high that you are one of the people who made this possible. I have learned many life lessons throughout this journey. Perhaps the biggest lesson I could pass to others is to simply be thankful for the gifts we receive and try to pass on the gifts we are capable of giving. Hard work, persistence, and patience are important, but we would not go far without the gifts from the powers and people who have created and shaped us. How much has any of us personally contributed to the roads we travel on, the clothing we wear, the food we eat, the shelter we live in, or even the book we just returned to the library?

Over the years, I have listed many people and many fortunate events in my gratitude journals. I am certain I could double the size of this dissertation if I were to list all of this here. Yet, in the interest of convention and keeping your patience, I will do my best to keep this section reasonable. I am extremely thankful for the unlimited guidance and patience from my committee members:

Drs. Andrew S. Davis, Eric E. Pierson, W. Holmes Finch, and Amy J. Harden. I am very thankful for every professor, teacher, and every student who has shaped me along the way. I have been a part of wonderful academic communities at Zion Lutheran School, Harbor Beach Community Schools, Huron Area Technical Center, Saginaw Valley State University, Ball State University, and Munroe-Meyer Institute. I am grateful for having a network of supportive friends. Some of those who have helped the most in giving me the pushes I have needed include: Darryl Booms, Justin Boseck, Gunnar Ingolfsson, Joe Nitz, and Eric Romzek. I cannot thank my own family members enough for all of the support and patience. I am thankful for my parents, Arnold and Shirley, for my siblings, Chad, Adam, and Ellie, and for my wife, Kiem. I also will never forget my loving grandparents, extended family, and all my friends and family from past and future generations. I have had an incredible opportunity to "stand on the shoulders of giants" and I hope that life presents me with many opportunities to help others ascend even higher.

Table of Contents

Chapter I

Introduction

Many definitions of the word "malingering" can be found in the scientific literature. Modern definitions tend to define malingering as a practice of intentionally modifying or misrepresenting one's own behavior for purposes of feigning or fabricating symptoms associated with a particular disorder or disease; individuals engaging in this behavior do so to avoid an external punishment or acquire an external incentive (Slick, Tan, Sherman, & Strauss, 2011; American Psychiatric Association, 2013). Most of the current definitions of malingering have relatively subtle differences in wording. Slick, Tan, Sherman, & Strauss (2011) have defined malingering as:

> The exaggeration and/or fabrication of deficits in malingering is a volitional behavior directed toward a substantial external incentive, either the acquisition of something desired or the escape from an undesirable duty, obligation, or punishment (p. 460).

The American Psychiatric Association in the *Diagnostic and Statistical Manual of Mental Disorders – 5th Edition* (*DSM-5*; American Psychiatric Association, 2013) have defined malingering as:

> The essential feature of malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives such as avoiding

> military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs.

The wording in the two prevalent definitions presented above demonstrate a large agreement in defining malingering; however, research is still exploring the possibilities of malingering being a mental disorder. The American Psychiatric Association in the *DSM-5* (2013) noted that malingering is not a mental disorder; however, some researchers (e.g. Raine, 2003) have argued that malingering fits the criteria of a mental disorder better than some of the other recognized mental disorders.

Interest in malingering has been increasing in recent years. Carone and Bush (2013) have suggested that scholarly attention to malingering has substantially increased in the last two decades. In clinical neuropsychology, the beginning of this scholarly attention to malingering appears to have been set in motion in 1978 with Heaton, Smith, Lehman, and Vogt's article, "Prospects for faking believable deficits on neuropsychological testing." In this article, Heaton and colleagues (1978) provided evidence that neuropsychologists' clinical intuition alone may be insufficient for detecting malingering. In the study, it was found that neuropsychologists' performance in classifying simulated malingerers was only marginally better than chance.

Rationale of the Study

Historically, it is clear that malingering has been present for aeons. Developments throughout human history, particularly the increasing availability of welfare support since the 1880's (see Wessely, 2003), has ushered in the culmination of interest that has been demonstrated in the neuropsychological literature in recent decades. Currently, all psychologists and neuropsychologists who are cognizant of current law and ethical code have some level of interest in topics relating to symptom validity and malingering as all psychologists are required to strive for accuracy throughout the evaluative process. The American Psychological Association noted in standard 9.01 of the *Ethical Principles of Psychologists and Code of Conduct*:

> Psychologists base opinions contained in their recommendations, reports and diagnostic or evaluative statements, including forensic testimony, on information and techniques sufficient to substantiate their findings (American Psychological Association, 2010, para. 1).

Clinicians engaging in neuropsychological evaluation may have an even greater onus to produce accurate results as these evaluations may result in significant financial gains or losses. Toomey, Kucharski, and Duncan (2009) note that in legal proceeding, the results of these evaluations could impede necessary treatment, influence the sentencing of an individual, or result in a defendant being acquitted by reason of insanity.

While neuropsychologists following the ethical code of the American Psychological Association strive for accuracy in their work, it is clear that not all neuropsychologists routinely employ standalone measures specifically designed to assess symptom validity. For example, Sharland and Gfeller (2007) found that a little more than half of a sample of National Academy of Neuropsychology professional members (n = 712) frequently used measures of symptom validity during their evaluations. For neuropsychologists who are involved in addressing financial compensation claims or personal injury litigation, about 79% employed tests of symptom validity (Slick, Tan, Strauss, and Hultsch, 2004). Standalone measures of symptom validity may not be employed for a variety of reasons. In some cases, these measures may not be deemed necessary as there is no apparent noteworthy incentive for examinees to provide suboptimal effort. In some cases, cost or availability of these measures may be prohibitive. O'Bryant, Engel, Kleiner, Vasterling, and Black (2007) noted that these standalone measures of symptom validity are underutilized in part because of the significant administration time required.

Because of the administration time required for some standalone measures of symptom validity, it is of interest to find ways to assess symptom validity with minimal time commitment. A number of

researchers have attempted to look for patterns in commonly administered psychological and neuropsychological assessments in hopes of finding objective and reliable ways of assessing symptom validity. Finding these "embedded validity indicators" in commonly used assessments is a very important task as practitioners may be able to efficiently and accurately measure symptom validity via information that may be readily available in a routine neuropsychological examination (Novitski, Steele, Karantzoulis, & Randolph, 2012). The potential increases in evaluation accuracy and the sheer time-savings available by creating and understanding embedded validity indicators provide a strong rationale for continuing studies in this area.

Many individuals who malinger claim to have impaired memory and it has been recommended that clinicians consider using tests that appear to measure memory for the purposes of assessing symptom validity (Lezak, Howieson, & Loring, 2004). The current study explored how performance on one of the newest widely used measures of memory, the *Wechsler Memory Scale, Fourth Edition* (*WMS-IV*; Wechsler, 2009) related to some of the most widely known standalone measures used for assessing symptom validity: the *Test of Memory Malingering* (*TOMM*; Tombaugh, 1996), and the *Word Memory Test* (*WMT*; Green, 2003). Part of the rationale of the study was to add to the research base on how current measures of memory and symptom validity compare. In standard 9.01 of the American Psychological Association's code of ethics (defined in a prior paragraph), it is noted that psychologists need to base their conclusions on data from techniques that provide the necessary data. The current study provided statistical data on how measured memory performance relates to an examinee's measured symptom validity; it also provided data on how one of the most recent validated measures of memory, the *WMS-IV*, related to some of the most commonly used measures for assessing symptom validity, the *TOMM* and *WMT*.

A related rationale for the current study could be determining if subtests from the *WMS-IV* and the *WMT* and *TOMM* are in fact measuring the same construct. If the relationship between

performance on subtests of the *WMS-IV* and the *WMT* and *TOMM* relate too strongly, a jangle fallacy may be identified. The jangle fallacy, as described by Kelley (1927), is a situation where multiple labels are used to describe the same underlying construct.

Significance of the Study

The study explored the utility of using performance patterns in a commonly utilized measure of memory, the information obtained from assessment with the *WMS-IV*, to compare performance to standalone measures of symptom validity, the *TOMM* and *WMT*. Understanding the relationship between these measures could provide a significant contribution to the field of neuropsychology as practitioners may be better informed as to how examinees' performance on a commonly used memory measure could relate to performance on standalone measures of symptom validity. This knowledge could save practitioners valuable evaluation time and improve the accuracy of evaluation conclusions by providing insight into the necessity of using additional standalone measures. Establishing relationships between performance on the *WMS-IV*, *TOMM*, and *WMT*, could provide a useful segue to the establishment of practical validity indicators.

The study explored what standalone measures may be more useful or appropriate when using the *WMS-IV*. The *WMS-IV* contains subtests that measure visual memory and subtests that specifically measure auditory memory. Standalone measures of symptom validity such as the *WMT* and *TOMM* contain tasks that appear to measure visual or auditory memory. With this information in mind, an exploration of how the visual and auditory memory subtests of the *WMS-IV* relates to the *WMT* and *TOMM* could reveal interesting relationships. Knowledge of these relationships could be useful to neuropsychologists as it would imply that suspect performance on certain parts of the *WMS-IV* could warrant usage of a specific standalone symptom validity test.

Study Procedures

The study used data which was obtained from a larger neuropsychological dataset. All data was obtained from college students participating in a psychology research pool. Data utilized in the study was from standard administrations of the *TOMM, WMT,* and *WMS-IV*.

Research Questions

R_1 How much variance in the WMS-IV subtests does the TOMM account for?

R_2 How much variance in the WMS-IV subtests does the WMT account for?

R_3 How much variance in the WMS-IV composites does the TOMM account for?

R_4 How much variance in the WMS-IV composites does the WMT account for?

R_5 Does the examination of the means of all variables in the obtained sample look comparable to the means found in the literature?

R_6 How much correlation is present between TOMM and WMT scores?

R_7 How much correlation is present between WMS-IV, TOMM, and WMT scores?

Limitations of the Study

One of the largest limitations to the current study is related to the nature of the sample used. By using a convenience sample of college student who were predominantly healthy young adults without substantial incentive to give full effort or reduced effort, it is unknown how the obtained data for the current study would generalize to clinical populations suspected of malingering. The sample of students who participated in the current study differed from many clinical populations on demographic factors such as age, gender, and level of education. The sample of students who participated in the study was obtained from a research participant pool from one university in the Midwest. Participants were offered research participation credit for an undergraduate psychology class; however, participants were allowed to withdraw and still receive participation credit.

During the study, participants were asked to give full effort although there was no incentive or penalty related to the actual amount of effort given. With data obtained in the current study, it cannot be determined whether or not the relationships amongst the variables in the *WMT*, *TOMM*, and *WMS-IV* would differ for participants engaged in simulated malingering or examinees engaged in actual malingering. Elucidating relationships between measures were complicated by the fact that the range of performance from the sample resulted in very little variability in symptom validity test scores. Little variability in symptom validity scores impeded the establishment of relationships between the measures used in the study.

Delimitations of the Current Study

The current study utilized some of the most highly regarded measures of memory and symptom validity. The *WMS-IV*, *TOMM*, and *WMT* have either existed for over a decade or have had earlier versions in existence for over a decade. The sheer number of studies and practitioners who routinely use these measures in practice could attest to the quality of these instruments.

Unlike in field studies, the circumstances for taking neuropsychological assessments in the current study are uniform. All participants in the current study were engaged in the assessments for the purposes of fulfilling research participation requirements related to a class. These participants very likely had uniform incentive for participation and presumably put forth similar levels of effort. The consistency in the circumstances for assessment administration and incentive suggested that the findings are reliable and replicable.

List of Terms

Amnestic Mild Cognitive Impairment (aMCI): Amnestic Mild Cognitive Impairment (aMCI) consists of memory complaints, isolated memory dysfunction, intact cognitive functioning, adequate activities of daily living, and no significant evidence of dementia (Mariani, Monastero, & Mecocci, 2007).

Coaching: In the context of neuropsychological assessment, coaching is a type of counseling that may encourage exaggeration of symptoms, feigning of symptoms, or resistance and defensiveness during the assessment. Coaching can include test coaching, symptom coaching, or a combination of coaching (Powell, Gfeller, Hendricks, & Sharland, 2004).

Cut Score / Cutting Score / Cut-off score: In the context of symptom validity tests, a cut score is a performance cut-off score that can provide an examiner with data on the likelihood of the examinee providing sufficient on a test. Cut scores are designed to minimize false positive diagnoses for examinees providing legitimate effort while accurately detecting low effort as much as possible (Larrabee, 2003).

Dementia: Dementia often refers to degenerative dementias (e.g. Alzheimer's Disease) that often impact older populations although it could refer to other cognitive declines by some definitions; Dementia is considered to be a *major neurocognitive disorder* (*DSM-5*; American Psychiatric Association, 2013).

Embedded Measures: Embedded measures, sometimes called embedded validity indicators, allow examiners to assess symptom validity based upon information obtained from subtests already administered throughout the course of a standard neuropsychological evaluation. Embedded measures are often based upon subtests that appear to be sensitive to brain injury; however, the subtests are relatively resistant to brain dysfunction and can therefore be used to assess symptom validity (Novitski, Steele, Karantzoulis, & Randolph, 2012).

External Incentive: In the context of neuropsychological evaluation, an examinee may choose to malinger to obtain medications, receive financial compensation, avoid criminal sentences, evade military duties, or benefit in some other way. These reasons for malingering are often referred to as external incentives (McDermott, Leamon, Feldman, & Scott, 2009).

Factitious Disorder: Individuals with factitious disorder are characterized by intentionally fabricating or feigning symptoms solely for the purposes of being identified as ill or as a patient (McDermott, Leamon, Feldman, & Scott, 2009).

False Negative: A false negative or a false negative error in the context of symptom validity testing, is an error that occurs when a bona fide malingering examinee is not detected (Faust & Ziskin, 2011).

False Positive: A false positive or a false positive error in the context of symptom validity testing, is an error that occurs when an examinee gives full effort yet is inaccurately classified as a malingerer (Faust & Ziskin, 2011).

Fixed Test Battery: Neuropsychologists who employ a fixed test battery administer the same tests to all examinees regardless of the referral question or other circumstances. Using a fixed test battery has the advantage of allowing examiners to compare the performance of an examinee to the performance of a large number of people who have taken the same tests (Holtz, 2011).

Flexible Test Battery: Neuropsychologists who employ a flexible test battery administer tests based upon the examinee's referral questions or unique needs. The advantage of using a flexible test battery is that by tailoring test batteries to the examinee, time usage and expense can be minimized and additional tests can be administered to explore areas of interest (Holtz, 2011).

Forced Choice Test: A forced choice test presents examinees with two response choices (or another limited number of choices) in which only one item is correct. In the context of assessing symptom validity, it may be of particular interest if an examinee scores below chance or exhibits other signs of reduced symptom validity (Frederick & Speed, 2007).

Functional Magnetic Resonance Imaging (fMRI): Functional magnetic resonance imaging (fMRI) is a type of imaging that can provide three-dimensional images of brain and show changes in blood flow that tend to be correlated with mental functioning (functional magnetic resonance imaging; *The American Heritage Medical Dictionary*, 2007).

Intelligence: Intelligence is a psychological construct with definitions varying based upon schools of thought. Many definitions of intelligence resemble a definition set forth by David Wechsler. Wechsler defined intelligence as "a global capacity to interact with one's environment through purposeful action and rational thought (Clauss-Ehlers, 2008)."

Intelligence Quotient (IQ): Originally, in the early 1900's, an intelligence quotient was an intelligence test scoring mechanism that was based upon a ratio of measured mental age to chronological age. From the mid 1900's onwards, an intelligence quotient was commonly computed based upon an individual's performance compared to the average performance of a peer group (Clauss-Ehlers, 2008).

Litigation: A litigation is a legal dispute or a judicial contest that is brought to a court for the purposes of enforcement of one's legal rights (Lehman & Phelps, 2005).

Malingered Neurocognitive Dysfunction (MND): Malingered Neurocognitive Dysfunction (MND) is a voluntary exaggeration or feigning of cognitive dysfunction for the purposes of receiving substantial compensation or other gain, avoiding duties, or evading responsibilities. It can be identified based upon criteria set forth by Slick, Sherman, and Iverson (1999). These criteria define response bias patterns that would constitute MND, symptom exaggeration or misreporting, and rule-out criteria for MND based upon other disorders or abnormalities that could account for the observed performance of the examinee (Larrabee, 2007)

Malingering: Malingering is the practice of intentionally modifying or misrepresenting one's behavior for the purposes of fabricating or feigning symptoms of a disorder or disease. Individuals who engage in malingering do so for the purposes of acquiring an external incentive or avoiding an external punishment (Slick, Tan, Sherman, & Strauss, 2011; American Psychiatric Association, 2013).

Mild Cognitive Impairment (MCI): Mild Cognitive Impairment (MCI) consists of deficits in cognitive abilities without significant impairment in activities of daily living. MCI is often associated with aging populations and may sometimes be a precursor to Alzheimer's Dementia. Several subtypes

of MCI have been proposed, yet diagnostic criteria has not been universally accepted (Mariani, Monastero, & Mecocci, 2007).

Mild Traumatic Brain Injury (mTBI): Mild Traumatic Brain Injury (mTBI) is a caused by a head injury that is significant enough to disrupt the normal functioning of the brain. Individuals with mTBI have at least one of the following: a loss of consciousness of less than 30 minutes, a disruption of mental state at the time of injury, a focal neurological deficit, a Glasgow Coma Scale score between 13 and 15, or an amnesia after the injury that lasts less than 24 hours (Messé, Caplain, Pélégrini-Issac, Blancho, Lévy, Aghakhani, & ... Lehéricy, 2013).

Neuropsychologist: A neuropsychologist, or clinical neuropsychology, engages in assessment, diagnosis, and treatment of individuals with brain injury or illness. The neuropsychologist relies on knowledge of the brain for the purposes of helping people with brain impairment (Holtz, 2011).

Neuropsychology: Neuropsychology, or clinical neuropsychology, is an applied psychology specialty that examines the behavior of individuals with normal brain functioning and determines how brain structure and function differs for individuals suffering from brain injury or disease (Holtz, 2011).

Sensitivity: Sensitivity of measures designed to measure symptom validity or response bias is related to how effectively the measure detects malingerers. Measures with high sensitivity tend to have a low false negative rate (Faust & Ziskin, 2011).

Simulator: In the context of the malingering literature, simulators are typically healthy nonclinical research participants who are asked to feign deficits related to head injuries or psychopathology. It is often assumed that the performance of simulators, like the performance of bona fide malingerers, will be lower or in some way different from individuals with legitimate head injuries or illness (Ju & Varney, 2000).

Specificity: Specificity of measures designed to measure symptom validity or response bias is related to how effectively the measure correctly categorizes individuals giving full effort as being non-

malingerers. Measures with high specificity tend to have a low false positive rate (Faust & Ziskin, 2011).

Structured Interview: Structured interviews are generally created to reduce missed or inaccurate diagnoses by standardizing the questions asked during an interview, properly sequencing these questions, and quantifying responses (Rogers, 2001).

Symptom Validity Test (SVT): SVTs are measures that are designed to provide objective information regarding the level of effort that a patient is giving during testing. SVTs are generally designed to be insensitive to genuine cognitive or neurological deficits (Powell, Gfeller, Hendricks, & Sharland, 2004). SVT can take the form of either measures embedded into standard neuropsychological tests or free-standing tests (Henry, Heilbronner, Mittenberg, Enders, Stevens, & Dux, 2011).

Chapter II

Review of the Literature

Although numerous definitions of the word "malingering" can be found in the literature, many of the definitions have components in common. Definitions tend to suggest that individuals who malinger, in contrast to individuals experiencing some other conditions of clinical attention, are intentionally modifying or misrepresenting their behavior or capabilities in certain environments or contexts. Definitions, such as those provided by Slick, Tan, Sherman, & Strauss (2011), the authors of the *Diagnostic and Statistical Manual of Mental Disorders – 4th Edition Text Revision* (*DSM-IV-TR*;American Psychiatric Association, 2000), and the authors of the *Diagnostic and Statistical Manual of Mental Disorders – 5th Edition* (*DSM-5*; American Psychiatric Association, 2013) also tend to suggest that individuals who engage in malingering attempt to fabricate or feign symptoms of a disorder or disease. Finally, many of the definitions of "malingering" propose that individuals who engage in malingering are doing so for the purposes of acquiring an external incentive or avoiding an external punishment (e.g. Slick et al., 2011; American Psychiatric Association, 2013).

Behaviors that constitute malingering have been present for a long time. According to Wessely (2003), malingering has existed in some form for as long as humans have gathered together into societies with responsibilities for individual members. For example, malingering in the military has been documented since at least the Trojan War (Palmer, 2003). The early roots of the word

"malingering," have been said to have originated with the military in unknown ancient history; the actual word, "malingering" originated from the French (Palmer, 2003). The first known use of the French word "malinger" was in 1820 (malinger; Merriam-Webster.com n.d).

In the context of medical evaluations, malingering has likely been present for as long as this behavior had incentive. Malingering in medical evaluations appeared to become more common as industrialized countries introduced laws that would provide support for those who could not work. Under Imperial Chancellor Bismarck of Germany, the 1883 Sickness Insurance Act, the 1884 Accident Insurance Law, and the 1889 Old Age and Disability Insurance Acts were instated (Wessely, 2003). During these years, the incentives for malingering increased. In Great Britain, the Workmen's Compensation Act of 1908 and the Lloyd George National Insurance Act of 1911 provided similar incentives to feign illness or injury (Wessely, 2003). Interestingly, it seems to have taken decades for more universal, agreed-upon definitions to appear although there are still some differences in the use of the term. In psychiatry, medicine, and related fields, more formal definitions of malingering appeared in the middle of the 20th century. The sixth edition of *International Classification of Diseases* (*ICD-6*, World Health Organization, 1948) provided a code specifically for "malingering" whereas a code for malingering was not available in earlier versions. According to Carone and Bush (2013), in the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM*; American Psychiatric Association, 1952) malingering appeared as a term in the index with no set criteria. The *Diagnostic and Statistical Manual of Mental Disorders – 2nd Edition* (*DSM-II*; American Psychiatric Association, 1968), defined malingering as being a "conscious behavior" that must be differentiated from "hysterical neurosis, conversion type" (Carone & Bush, 2013, p. 8). The *Diagnostic and Statistical Manual of Mental Disorders – 3rd Edition* (*DSM-III*; American Psychiatric Association, 1980), provided a definition for malingering that read:

> The essential feature is the voluntary production and presentation of false or grossly exaggerated physical or psychological symptoms. The symptoms are produced in pursuit of a goal that is obviously recognizable with an understanding of the individual's circumstances rather than of his or her individual psychology (p. 331).

The *DSM-III* provided examples of reasons for why someone might malinger, such as avoiding military duty, and noted that malingering must be differentiated from conversion disorders and factitious disorders. The *DSM-III* also acknowledged that malingering can be adaptive in situations such as when one has been captured by the enemy during wartime. The *Diagnostic and Statistical Manual of Mental Disorders- 3rd Edition Revised* (*DSM-III-R;* American Psychiatric Association, 1987) retained an explanation of malingering that was nearly identical to the *DSM-III*, but the wording "voluntary production and presentation" was changed to "intentional production and presentation" and also the words "external incentive" replaced the word "goal" (p. 360). The *Diagnostic and Statistical Manual of Mental Disorders- 4th Edition* (*DSM-IV* ;American Psychiatric Association, 1994) featured a definition of malingering that was nearly identical to the definition in the *DSM-III-R*, but with the *Diagnostic and Statistical Manual of Mental Disorders – 4th Edition Text Revision* (*DSM-IV-TR*; American Psychiatric Association, 2000) the words "intentional production" was changed to "deliberate fabrication" (p.739). The wording of the definition of malingering in the *DSM-IV-TR* has very likely influenced other definitions that have appeared in the literature in more recent years. Current researchers in the field, such as Slick, Tan, Sherman, & Strauss (2011), have set forth similar definitions for the construct of malingering:

> The exaggeration and/or fabrication of deficits in malingering is a volitional behavior directed toward a substantial external incentive, either the acquisition of something desired or the escape from an undesirable duty, obligation, or punishment (p. 460).

As noted by Bush, Ruff, Tröster, Barth, Koffler, Pliskin, Reynolds, and Silver (2005) neuropsychologists are particularly concerned with the idea of malingering as they must be able to confidently determine the validity of their findings during evaluation and attribute their data to bona fide examinee impairment or reduced effort. A National Academy of Neuropsychology (NAN) position paper published in 2005 by Bush and colleagues defined malingering in a way that resembled the definitions of malingering present in the editions of the *Diagnostic and Statistical Manual of Mental Disorders* published prior to the *Diagnostic and Statistical Manual of Mental Disorders – 4th Edition Text Revision* (*DSM-IV-TR*; American Psychiatric Association, 2000). The NAN definition from 2005 retained the "intentional production" wording found in earlier versions of the *Diagnostic and Statistical Manual of Mental Disorders* while making the rest of the definition consistent with the definition found in the *DSM-IV-TR* (Bush et al., 2005, p. 420). The similarity of the definition provided in the NAN position paper with that provided in the *DSM-IV-TR* (American Psychiatric Association, 2000) suggests that the field of neuropsychology is largely in agreement with fields such as psychiatry on the definition of malingering.

There appears to be a general consensus that malingering is not a disease; however, there is debate over whether or not malingering should be considered a mental disorder. The designation of malingering set forth in the *DSM-IV-TR* (American Psychiatric Association, 2000) as well as the general consensus of important figures in malingering research agree that malingering is not a pathological or disease process (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). While the *Diagnostic and Statistical Manual of Mental Disorders – 5th Edition* (*DSM-5* ; American Psychiatric Association, 2013) notes that malingering is not a mental disorder, other experts in the field have argued or explored the possibility of malingering being a mental disorder. Historically, there have been a number of viewpoints on whether or not malingering was a mental disorder. During the World War II era, the popular viewpoint on malingering was that only ill people would attempt malingering

(Halligan, Bass, & Oakley, 2003). Raine (2003) compared criminal behavior, malingering, and conditions listed in the *DSM-IV* against a number of criteria that could be used to define a mental disorder and concluded that malingering fit the criteria of mental disorder better than some other conditions listed in the *DSM-IV*. It could be argued that malingering could be more representative of a mental disorder than conditions such as caffeine intoxication and schizotypal personality disorder due to higher treatment-seeking behavior and higher deviation from social norms (Raine, 2003, p.102). The lack of agreement on whether or not malingering is a mental disorder suggests that definitions of malingering which assert a mental disorder status for this condition may be problematic more than other definitions. Furthermore, this ongoing debate reveals that more research and explorations of the constructs of malingering and mental disorders is needed.

Malingering shares a number of features with other conditions or disorders. With this in mind, there may be advantages and disadvantages to discussing malingering alongside similar conditions. The *DSM-5* departs from the previous versions of the manual by seemingly deemphasizing malingering as being a unique construct due to its similarities to factitious disorder. The *DSM-5* discusses malingering and factitious disorder under a single heading within a broader discussion of conversion disorder. The DSM-5 indicates that assessment of conscious intention is unreliable and is not necessary for the diagnosis of a conversion disorder; however, clear signs of feigning would warrant a diagnosis of factitious disorder or malingering depending upon whether an individual wanted to assume a sick role or wanted to gain a significant incentive. Discussing malingering alongside factitious disorder in the *DSM-5* likely reflects the observance of the similarities by other researchers in the field. Consistent with the discussion of malingering and factitious disorder in the *DSM*-5, Lande and Williams (2013) noted these can present in an almost identical way. The notable difference between malingering and factitious disorder is the incentive for the disorder; rather than seeking incentives or

avoiding undesired situations, people with factitious disorder merely wish to be considered ill (Lande & Williams, 2013).

Although there is some general agreement across disciplines as to what constitutes malingering, the historical differences do highlight some concerns in not having a universal definition. Current definitions of malingering do not tend to set clear boundaries of what can be considered malingering nor do these definitions make it clear whether or not malingering is a categorical or dimensional construct (Raine, 2003). For example, Raine (2003) posed a question of whether or not exaggerating a common cold to get a day off of work could qualify as malingering. In some sense, this would fulfill the criteria of some of the currently accepted definitions of malingering as an individual who exaggerates sickness associated with a common cold is potentially fabricating deficits in functioning and is doing so for the purposes of avoiding something that this individual deems as aversive. On the other hand, the magnitude of symptom exaggeration, the value of the external incentive, and the commonality of this behavior raises a question of whether or not identifying this behavior as malingering would be worthwhile from a categorical viewpoint. A related definitional problem of malingering and of factitious disorder is the necessity of determining how willful or conscious behaviors must be to align with either condition. Halligan, Bass, and Oakley (2003) pointed out that proper differentiation between factitious disorders and malingering require the possibly impossible task of determining the level of conscious awareness, the motivations of the examinee, and the "degree of consciously mediated intention" (p. 9).

Within the construct of malingering, subtypes of malingering have been defined. Resnick (1997) has defined three types of malingering: Pure malingering, partial malingering, and false imputation. Pure malingering is when an examinee attempts to completely fabricate symptoms of a specific psychiatric or neuropsychological condition. In partial malingering, an examinee may have a limited set or intensity of symptoms, yet that examinee attempts to exaggerate those symptoms. Some

individuals may indeed have genuine neuropsychological deficits, but may exaggerate the problem for personal gain (Iverson, 2003). Examinees engaging in false imputation misattribute their current symptoms to another event. For example, an examinee who has a lengthy history of memory or attention problems may claim that these problems were related to a recent car accident. Although current nosology (e.g. *DSM-5)* does not tend to require greater specificity when identifying "malingering", more specific terms are available for clinicians and researchers to utilize. In many situations, defining specific subtypes of malingering may be unnecessary, but using terms that are less precise, vague, or not validated could pose problems. Although other words could potentially be used in lieu of "malingering", it may be problematic to deviate from this word in some cases. Rogers and Neumann (2003) caution against using the term "secondary gain" in place of "malingering" as the authors state that the construct of "secondary gain" does not have empirical backing.

Malingering As a Problem in Assessment

Feigning and exaggerating medical and psychiatric problems have a very long history, but it has only been within the past two decades that increased scholarly attention has been devoted to understanding this phenomenon (Carone & Bush, 2013). For decades, neuropsychologists have realized that it is necessary to use objective measures to supplement clinical judgment in some cases. In 1978, an article, "Prospects for faking believable deficits on neuropsychological testing", specifically addressed the topic of malingering and the need to assess symptom validity in objective ways (Heaton, Smith, Lehman, & Vogt, 1978). During this study, participants with verified head injuries (n = 16) and participants who were simulating malingering (n = 16) were administered a battery of neuropsychological tests. In this study, the *Wechsler Adult Intelligence Scale* (*WAIS*; Wechsler, 1955), the *Minnesota Multiphasic Personality Inventory* (*MMPI;* Hathaway & McKinley, 1942) and the *Halstead-Reitan Neuropsychological Battery* (*HRNB*; Reitan & Wolfson, 1993) were administered by an experienced technician. This technician was "blinded" by being completely

unaware of the fact that the participants were not real patients. This study used neuropsychologists (*n* = 10) with a range of experiences, which would likely be representative of practicing neuropsychologists in the field, as judges. These neuropsychologist judges were asked to determine which participants were malingering by solely examining their test scores. The neuropsychologist judges were able to correctly classify between 50% and 68% of participants. With neuropsychologist judges having identification accuracy that was barely better than chance, this study strongly suggested that examination of test scores of common measures alone is not sufficient for the identification of malingerers. Heaton and colleagues (1978) did, however, note that these neuropsychologist judges may have had limited experience with malingerers and that these judges were deprived the chance to directly observe patients or know the observations of the technicians. So, the accuracy rates could have improved if the neuropsychologists had access to data that they would often be able to obtain in routine practice.

Avoiding over and under-identifying malingering is crucial and often a key component of the assessment process. For example, to illustrate the high-stakes nature of this process, it is possible that misclassifying individuals could result in acquitting a defendant by reason of insanity, altering the severity of a sentence, finding an individual incompetent to stand trial, or impeding necessary treatment (Toomey, Kucharski, & Duncan, 2009). Interestingly, despite the very high stakes associated with misclassification of malingering, the American Psychological Association has not made specific ethical principles regarding malingering in the *Ethical Principles of Psychologists and Code of Conduct*; however, in standard 9.01, it generally addresses malingering by noting:

> Psychologists base opinions contained in their recommendations, reports and diagnostic or evaluative statements, including forensic testimony, on information and techniques sufficient to substantiate their findings (American Psychological Association, 2010, para. 1).

Methods of Malingering Assessment

With the high-stakes nature of many neuropsychological evaluations in mind, it is crucial for neuropsychologists to adopt measures and methods that elucidate which patients are giving genuine effort. There are presently a number of methods of detecting malingering available to neuropsychologists. Practitioners can attempt to detect malingering through clinical intuition which could be informed by observations of the examinee and review of the examinee's history. Practitioners can also use specialized measures, such as symptom validity tests to directly address the possibility of malingering during the course of an examination. Practitioners can also examine performance on commonly used neuropsychological assessment measures and use formulas to determine likelihood of patterns of performance. Regardless of the methods chosen to detect malingering, there are two general strategies used for the detection of malingering: finding examinees to be excessively impaired and finding examinees to exhibit unexpected performance patterns (Rogers & Bender, 2003).

Some guidelines have been proposed for how neuropsychologists should attempt to examine the possibility of malingering. To detect malingering, Iverson (2003) recommends that practitioners use specialized symptom validity tests as well as examine the performance patterns on common neuropsychological tests that are administered for the purposes of assessing other areas of functioning. It is also recommended that clinicians intersperse validity indicators throughout the evaluation (Iverson, 2003). Neuropsychologists who opt to only examine validity indicators during certain portions of an evaluation risk not detecting possible fluctuating effort given by examinees. Heilbronner and colleagues (2009) also offered a set of guidelines for evaluating neuropsychological response validity and communicating those findings. Among these guidelines, it was noted that neuropsychologists use psychometric indicators to ensure validity, use stand-alone and embedded validity indicators, compare test data with known "real-world" performance, compare obtained data

with documented disorders, intersperse validity indicators throughout the assessment, and use serial evaluations when possible.

Using measures beyond clinical intuition can be crucial regardless of setting and incentive. Pella and colleagues (2012) using a sample of self-referred students from a Southern university psychological clinic (n = 986) and a control student sample (n = 182) concluded that university students who exaggerate symptoms may obtain their desired outcomes when symptom validity indicators such as empirically supported embedded validity indicators are not used. In this study, some self-referred students from the study with Pella and colleagues (2012) had known external incentives while others had no known external incentives. Although flexible test batteries were used, all of the examinees had *Wechsler Adult Intelligence Scale – Third Edition* (*WAIS-III*; Wechsler, 1997a) and *Wechsler Memory Scale – Third Edition* (*WMS-III*; Wechsler, 1997) data and some examinees had *Personality Assessment Inventory* (*PAI*; Morey, 1991) data. Symptom validity in this study was assessed via embedded measures on the measures of memory and cognitive functioning. Pella and colleagues (2012) noted that failure of embedded measures was high for students with and without external incentive, but found the expected patterns of students with external incentive having poorer neuropsychological performance and higher rates of failing embedded measures. Undoubtedly, individuals in a college setting as well as many other settings may be capable of malingering and remaining undetected if symptom validity indicators are not used.

Unfortunately, several surveys have demonstrated that symptom validity tests are sometimes not used even in settings where malingering would be likely. Slick, Tan, Strauss, and Hultsch (2004) found that the majority (79%) of neuropsychologists who handle financial compensation claim or personal injury litigation use a symptom validity test; however, this figure demonstrates that a sizable minority of neuropsychologists in this environment do not employ these tests. More recently Sharland and Gfeller (2007), using a survey from a sample of NAN professional members (n = 712), found that a

little more than half of the members frequently utilized measures of symptom validity within their evaluations. Sharland and Gfeller (2007) also found that there was inconsistency regarding whether or not practitioners warned their patients about the possibility of symptom validity tests being used; 52% rarely or never give a warning while 27% often or always provide a warning. It also appears that practitioners have their own preferences regarding measures and therefore employ a wide variety of measures. Sharland and Gfeller (2007) found that the five most popular symptom validity or response bias tests were the Test of Memory Malingering (*TOMM*; Tombaugh, 1996), two scales from the Minnesota Multiphasic Personality Inventory, Second Edition (MMPI-2; Butcher et al., 2001), the *Rey 15-Item* test (*Rey*-15; Rey, 1964), and the *California Verbal Learning Test* (*CVLT*; Delis, Kramer, Kaplan, & Ober, 1987).

Lezak, Howieson, and Loring (2004) noted that people who attempt to malinger often claim to have impaired memory. With impaired memory being a common complaint for malingerers, it has been recommended that clinicians consider using malingering or symptom validity tests that appear to measure memory (Lezak et al., 2004). Lezak and colleagues (2004) have also recommended that clinicians look for patterns in existing memory measures.

Unfortunately, attempting to assess symptom validity in one area of functioning does not always result in an accurate assessment of symptom validity in another area. Haggerty, Frazier, Busch and Naugle (2007) investigated how well separate measures of cognitive symptom exaggeration relate to each other and well as how these cognitive symptom exaggeration measures relate to psychopathological symptom exaggeration. To address their research questions, Haggerty and colleagues (2007) used archival data from patients (n = 300) with neurological and neurodegenerative disorders referred for neuropsychological evaluation for assessment of functioning. Patients in the sample used for this study appeared to match the expected demographic of patients at a Midwestern hospital. About 16% of the sample was seeking compensation for their condition or were involved in

litigation. All selected patients completed the *VSVT* and *PAI* in a manner consistent with standard administration and interpretation. The performance of the patients was classified as valid or invalid based upon scores obtained on the *VSVT*. Haggerty and colleagues (2007) found moderate correlations between *VSVT* accuracy and latency scores (r = -.37 to -.64). Modest correlations of less than r = .17 were found amongst the *VSVT* accuracy and latency scores and the *PAI* Negative Impression Management validity scales; however, no significant correlations with the other *PAI* validity scales were found. This study, although exploratory in nature, suggests that validity scales within a measure may have notable relationships; however, validity scales across measures which assess different constructs may tend to have weaker relationships. On a practical level, this also advocates for the use of symptom validity measures for multiple areas of functioning during the course of an evaluation.

Assessing Malingering With Observational Techniques

Practitioners and researchers can sometimes detect malingering via careful observation and review of an examinee's history. Bush and colleagues (2005) indicated that malingering may be present if an examinee's self-reported history differs from documented history, the self-reported symptoms are inconsistent with known patterns of brain functioning, the self-reported symptoms are inconsistent with patterns on administered psychological tests, self-reported symptoms are inconsistent with behavioral observations, or the self-reported symptoms conflict with information gathered from other informants.

Aside from inconsistent historical information and questionable psychological test performance, sometimes neuropsychologists or other practitioners can detect malingering through physical examination. Descriptions of physical examination techniques for the detection of malingering have existed for decades. Waddell, McCulloch, Kummel, and Venner (1980) released a noteworthy publication that detailed some of the techniques available for differentiating between pain with physical correlates and "nonorganic physical signs." Waddell and colleagues (1980) wrote that finding three or

more categories of the five listed categories of physical signs is clinically significant. The five categories given were related to tenderness inconsistent with known neuroanatomy, back pain reported with certain loading and rotation tasks that should not produce back pain, differing abilities to move while being distracted, disturbances present over regions of the body that would be inconsistent with known neuroanatomy, or overreactions to physical tests. These "nonorganic physical signs", sometimes called "Waddell signs", may be helpful in identifying malingering; however, some researchers such as Fishbain, Cutler, Rosomoff, and Rosomoff (2004) have found little or no association between Waddell signs and four other methods of identifying malingering. Specifically, Fishbain and colleagues (2004) reviewed the findings and the scientific quality of 16 studies related to Waddell signs and found that 75% of the reports found no association between Waddell signs and 4 other methods of identifying malingering patients. This research is a reiteration of the common recommendation present in the literature: Multiple methods should be used when assessing the possibility of malingering.

Other physical examination techniques aside from those defined by Waddell and colleagues (1980) have been defined. Greer, Chambliss, and Mackler (2005) provided explanations of some other physical examination techniques for detecting malingering. Greer and colleagues (2005) made reference to a number of tests available for the detection of malingering. These tests included "McBride's Test, Mankopf's Test, Waddell's Test, Hoover's Test, the Abductor Test, the Arm Drop Test, and the Midline Split Test" (p.720). Greer and colleagues (2005) noted that the listed physical examination techniques tended to have few or no published studies examining the efficacy of the techniques and noted that "no examination technique objectively proves malingering" (p. 719).

Assessing Malingering With Embedded Validity Indicators

Some researchers and practitioners have attempted to look at patterns in common neuropsychological and psychological measures for the purposes of developing objective ways to

detect malingering. The advantage of examining these embedded validity indicators is that this information is often available in routine neuropsychological examinations (Novitski, Steele, Karantzoulis, & Randolph, 2012). In general, adopters of this method reason that the more unusual the results are on various assessments, the more likely it is that malingering is present.

Embedded validity indicators are available for a number of assessments used during neuropsychological evaluations. Trueblood and Schmidt (1993) assembled cut scores for some *Halstead-Reitan Neuropsychological Battery* (*HRNB*; Reitan & Wolfson, 2004) tests (e.g. *Seashore Rhythm Test*) and other measures (e.g. finger agnosia errors) and found that with 3 cut scores reached, approximately 14% of the cases were false positives. This indicates that while using cut scores for common neuropsychological tests for the purposes of identifying malingering, malingerers will often be identified; however, a sizable minority of people giving genuine effort will be erroneously identified as malingerers. Mittenberg, Rotholc, Russell, and Heilbronner's (1996) accuracy of identifying malingerers using the *HRNB* was similar to the accuracy found in Trueblood and Schmidt's (1993) study. Mittenberg and colleagues (1996) were able to use *HRNB* performance to correctly identify 88.75% of the groups with 93.8% true negatives and 83.8% true positives. Mittenberg and colleagues (1996) reached these conclusions by administering the *HRNB* to normal volunteers who were instructed to simulate malingering (n = 40) and nonlitigating head trauma patients (n = 40). By analyzing the test results via MANOVA, Mittenberg and colleagues (1996) found that the simulated malingerers generally performed worse than head trauma patients on tests measuring sensory, motor, and attentional performance; head injury patients tended to perform worse on more complicated cognitive measures such as the Tactual Performance Test and the Trails B test.

Silverberg, Wertheimer, and Fichtenberg (2007) developed an embedded measure, known as the Effort Index (EI), for the *Repeatable Battery for the Assessment of Neuropsychological Status* (*RBANS*; Randolph, 1998) which reportedly had "good discriminability" when differentiating between

groups with cognitive dysfunction and groups with poor effort (p. 841). Novitski and colleagues (2012) developed an effort scale (ES) for the *RBANS* and found that it was able to discriminate between people with amnestic disorders and poor effort markedly better than the EI developed by Silverberg and colleagues (2007). Novitski and colleagues (2012) developed their ES scale by using existing *RBANS* data obtained from patients with mild traumatic brain injuries (mTBIs; $n = 25$), patients with amnestic mild cognitive impairment (MCI; $n = 15$) and patients with probable Alzheimer's disease ($n = 54$) who had no known somatoform tendencies or incentives to alter performance for secondary gain. Novitski and colleagues (2012) compared their obtained data to the normal population in the *RBANS* standardization sample ($n = 540$). Novitski and colleagues (2012) found that only 17% of people in the standardization group scored below a specific cut score on a composite formed by summing digit span and list recognition scores whereas 78% of the impaired sample in the study scored below this cutting score. Novitski and colleagues (2012) concluded that examinees with scores under this level should be regarded as likely either being impaired or giving reduced effort during testing. Although embedded measures such as the ES on the *RBANS* lack specificity and sensitivity compared to some standalone measures, the authors pointed out the utility of this embedded measure as it was developed using information that can be automatically collected during the course of a brief neuropsychological evaluation.

Constantinou and colleagues (2005) found that some subtests of the commonly used *Wechsler Adult Intelligence Scale – Revised* (*WAIS-R*; Wechsler, 1981) were sensitive to malingering. Williams (2011), like Constantinou and colleagues (2005), found that certain subtests on the *WAIS-R* were more or less sensitive to malingering. Specifically, it was found that Digit Span, Arithmetic, and all of the Performance subtests were impacted more than the verbal subtests when examinees malingered. For the Williams (2011) study, determination of malingering, for the purposes of measuring *WAIS-R* subtest sensitivity was determined by results on the Test of Memory Malingering (*TOMM*; Tombaugh,

1996). A part of the *WAIS-R* that is particularly useful for detecting malingering is Digit Span and an extrapolated scale, Reliable Digit Span (RDS); malingering classification accuracy for Digit Span has been established (Heinly, Greve, Bianchini, Love, & Brennan, 2005). Miele, Gunner, Lynch, and McCaffrey (2012) found RDS to be superior to all of the other tested embedded validity indicators. Miele and colleagues (2012) reached these conclusions by conducting an analysis of archival data from examinees (n = 50) who were examined by a neuropsychologist for medico-legal purposes and who mostly had claims related to mTBIs. The examinees took the Rey 15 (Rey, 1964), the *TOMM* (Tombaugh, 1996), the Victoria Symptom Validity Test (*VSVT*; Slick, Hopp, Strauss, & Thompson, 1997), and the *Word Memory Test* (*WMT*; Green, 2003). The examinees also provided validity data through embedded measures on the *WAIS-R* (Wechsler, 1981) and the *Halstead-Reitan Neuropsychological Battery* (*HRNB*; Reitan & Wolfson, 1993). Miele and colleagues (2012) used cut scores from the standalone symptom validity tests that were specified in the manual and used cut scores for the embedded measures present in the literature. Miele and colleagues (2012) noted that some embedded measures were tested with multiple cut scores mentioned in the literature. A total of 17 possible embedded validity scores were analyzed. Miele and colleagues (2012) utilized a logistic regression to determine the ability of the embedded validity indicators to predict group membership for people failing or not failing standalone symptom validity indicators. It was found that only four of the embedded validity indicators could significantly group examinees based upon symptom validity. The four embedded validity indicators that were successful at grouping examinees were Reliable Digit Span, Category Test (CT) Total Errors on subtest 7, Speech Sounds Errors, and Tactile Finger Recognition Total Errors. Miele and colleagues (2012) found that of all embedded validity indicators, RDS proved to be the most useful. RDS with a proper cut score could correctly classify 74% of examinees; however, it was noted that there was a false-positive rate of nearly 20% with RDS. In addition to detecting malingering, Digit Span and Reliable Digit Span performance can serve as

indicators of suboptimal effort for individuals with Attention Deficit-Hyperactivity Disorder and individuals with Learning Disorders (Harrison, Rosenblum, & Currie, 2010).

Binder, Villanueva, Howieson, and Moore (1993) found that the profile of performance on the *Rey Auditory Verbal Learning Test* (*RAVLT*; Lezak, 1983) could be noticeably altered when examinees engaged in malingering. Binder and colleagues (1993) suggested examinees scoring below a certain score on the *RAVLT* recognition task should be carefully assessed for malingering; however, poor recognition scores are not pathognomonic of malingering. Conditions such as Alzheimer's dementia, aphasia, and alexia could also cause poor *RAVLT* recognition scores (Binder et al., 1993). Green, Rohling, Lees-Haley, and Allen (2001) also similarly found that the level of severity of neurological illness or head injury does relate to scores obtained on symptom validity tests.

Armistead-Jehle and Buican (2013) designed a study for investigating the efficacy of the *Advanced Clinical Solutions* package (*ACS*; Pearson, 2009) in comparison to the stand-alone *Word Memory Test* (*WMT*; Green, 2003) for the measurement of symptom validity. The ACS relies upon information obtained from assessment with the *Wechsler Memory Scale, Fourth Edition* (*WMS-IV*; Wechsler, 2009) and the *Wechsler Adult Intelligence Scales-Fourth Edition* (*WAIS-IV*; Wechsler, 2008). Using embedded measures and additional subtests (e.g. Word Choice Test), the ACS provides information and norms that can allow for users to obtain estimates of pre-morbid intelligence, executive function, social perception, and symptom validity. Armistead-Jehle and Buican's (2013) study utilized data from U.S. military members (n=280) with mild TBI. The authors noted that the *WMT* and all parts of the neuropsychological battery were administered using standard instructions aside from the examiner remaining in the room while the *WMT* was administered. Armistead-Jehle and Buican (2013) compared performance for the ACS and *WMT* at various base rate levels and found that the ACS had high specificity but low sensitivity. The authors found that at the 10% base rate level for the ACS, there were 32% of cases where the *WMT* was failed, but the ACS was passed and only 0.4%

of cases where the ACS was failed but the *WMT* were passed. The authors found that 62% of people passed both the *WMT* and ACS and just 6% of people failed both measures. The authors found that similar trends were present at other ACS base rates, but concluded that better specificity and sensitivity could be achieved at base rates higher than recommended by ACS documentation. Armistead-Jehle and Buican (2013) also concluded that embedded measures, such as the ACS can provide useful information, but that other methods for detecting malingering, such as using stand-alone symptom validity tests should also be employed.

Accuracy of embedded validity indicators can be improved by examining multiple measures. For example, using the WAIS-R Digit Span (Wechsler, 1981) scaled score, Knox Cube Test (Stone & Wright, 1980) total score, and Recognition Memory Test (Warrington, 1984), Iverson and Franzen (1994) established a 98% classification accuracy for malingerers and non-malingers based upon performance of 100 university students, federal inmates, and patients. Meyer and Volbrecht (2003) found that by using cutoff scores on a battery of nine common neuropsychological measures (e.g. *WAIS-R*, Wechsler, 1981; Trail Making Test, Reitan & Wolfson, 1985); Animal Naming (Strauss, Sherman, & Spreen, 2006), it was possible to classify over 700 examinees correctly with a 0% false positive rate and a 17% false negative rate. This level of accuracy could be obtained by considering people as malingering if they exceeded two cutoff scores (Meyer & Volbrecht, 2003).

Although methods such as examining built-in validity indicators in existing measures for malingering yield useful information, there is a consensus among experts suggesting information beyond that obtained with embedded validity indicators should be used for a determination of examinee symptom validity. Lezak, Howieson, and Loring (2004) noted positive findings with embedded measures suggest that a problem is likely present, but negative findings do not rule out the possibility of a problem. Heilbronner and colleagues (2009) recommended that practitioners employee both stand-alone symptom validity measures and embedded validity indicators when assessing malingering.

Whitney (2013, p. 233) wrote that "embedded symptom validity measures are best not used in isolation." Practitioners, therefore, should augment this information with other method of detecting poor effort.

Using Dedicated Symptom Validity Testing for Detecting Malingering

Practitioners can elect to utilize tests specifically designed to assess effort and symptom validity. Although symptoms validity tests can take a variety of forms, a common format is the "forced-choice test". Clinicians sometimes utilize forced-choice tests when malingering is suspected (Lezak et al, 2004). These forced-choice tests work by having an examiner present items to an examinee and then requiring an examinee to respond by choosing an answer from specific presented items. In the interest of test security of these assessments, more specific details about these tests are intentionally omitted from this manuscript.

According to Lezak and colleagues (2004), some of the most commonly used forced choice measures include: *Forced Choice Test* (Hiscock & Hiscock, 1989), *Portland Digit Reconstruction Test*, *Amsterdam Short-Term Memory Test*, *Test of Memory Malingering* (*TOMM*; Tombaugh, 1996), *Validity Indicator Profile*, and *The Victoria Symptom Validity Test* (*VSVT*). The *Rey 15-Item Test* or *Rey I* (Rey-15; Rey, 1964) is commonly used by neuropsychologists handling financial compensation claims (Slick, Tan, Strauss, & Hultsch, 2004). Slick and colleagues (2004) point out that quite interestingly, the *Rey I* is used frequently despite evidence of its inefficacy.

Although a number of forced-choice tests are available, some tests clearly have larger bases of research and larger user bases. The *TOMM* appears to be the most widely used and studied test of cognitive underperformance (Jelicic, Ceunen, Peters, & Merckelbach, 2011). One of the attributes about the *TOMM* that likely interests researchers and clinicians is its accuracy. Gervais, Rohling, Green, and Ford (2004) have found that the *TOMM* has less than a 1% false positive rate. Even with

the high accuracy provided by the *TOMM*, researchers are still finding other ways to examine symptom validity tests, like the *TOMM*, in hopes of further improving its accuracy. Gunner, Miele, Lynch, and McCaffrey (2012), for example, applied criteria developed for the *Word Memory Test* (*WMT*, Green, 2003) to the *TOMM* and found that a newly developed Albany Consistency Index (Gunner, et al., 2012) provided greater sensitivity and specificity for the *TOMM*. Denning (2012) also examined the *TOMM* and found that administration of just the first trial of the *TOMM* was more effective in identifying poor effort compared to administration of the entire *TOMM* when using the *TOMM* to predict results of the *Medical Symptom Validity Test* (*MSVT; Green, 2004*). It was also found that by adjusting cutoff scores for the *TOMM*, based upon whether or not neurological dysfunction was suspected, accuracy of the *TOMM* could be improved further (Denning, 2012).

The usefulness of symptom validity tests, such as the *TOMM*, can be partly determined by its ability to be usable and accurate with a wide array of people. The accuracy of the *TOMM* appears to be very high even when working with populations with bona fide impairments although there are some notable exceptions. Hill, Laurie, Kennedy, and Malamut (2003), for example, found that performance on the *TOMM* was impacted by temporal lobe dysfunction; however, performance on this measure was not impacted enough to be below the cutoff score for low effort or malingering. Teichner and Wagner (2004) found normal and cognitively impaired elderly people could be correctly classified using the *TOMM*; however, misclassification rates for elderly individuals with dementia were high. Rees, Tombaugh, and Boulay (2001) found that severe depression and affective state did not significantly impact examinee performance on the *TOMM*. Brooks, Sherman, and Krol (2012) also found that the *TOMM* could be largely passed by children who have been considered to be low functioning.

Although the *TOMM* may be a common choice for practitioners requiring symptom validity testing, some researchers have found that other symptom validity tests may offer superior attributes. Armistead-Jehle and Gervais (2011) found that the *Nonverbal Medical Symptom Validity Test* (*NV-*

MSVT; Green, 2008) had a greater sensitivity to poor effort compared to the *TOMM* while maintaining acceptable false positive rates using the *MSVT* and *Word Memory Test* (*WMT*, Green, 2003) as external criteria. In Armistead-Jehle and Gervais's (2011) study, examinees in the Province of Alberta, Canada who were claiming disability not related to head injury (n = 345) were given the *NV-MSVT* and the *TOMM* among other tests in a neuropsychological battery. Although Armistead-Jehle and Gervais (2011) were unable to know for sure which examinees were malingering due to the nature of their clinical sample, they used the *MSVT* and the *WMT* as the criteria in the study to determine whether or not examinees malingered. In this study, there was a tendency for participants to pass the *TOMM* and fail the *NV-MSVT*. Using the *MSVT* and *WMT* criteria, the authors made determinations of whether the *TOMM* was producing false negatives or the *NV-MSVT* was producing false positives. The authors found that the *NV-MSVT* was twice as sensitive as the *TOMM* to poor effort and maintained acceptable levels of specificity.

Researchers and practitioners have good alternatives to the *TOMM* available. This is important for situations in which a second standalone symptom validity test would be indicated or in situations where different tasks or modalities would be more appropriate for specific examinees. The *Word Memory Test* (*WMT*, Green, 2003) is a test designed to assess verbal memory, learning and symptom validity (Flaro, Green, & Robertson, 2007). The *WMT* is regarded by some as being unique among the available symptom validity tests because of the number of validation studies conducted on clinical samples (Green, Lees-Haley, & Allen, 2003). Hartman (2002) described the *WMT* as being "robust" and having a "large normative base" as well as being supported by ongoing research after its publication. The *WMT* is a computerized symptom validity test that assesses an examinee's ability to recognize previously presented word pairs (Green, 2003). The *WMT* can also be administered via a paper and pencil method. Hoskins and colleagues (2010) have found that the computerized version of the *WMT* appears to be equivalent to the orally presented version. Although the *WMT* and *TOMM* have

similar goals of measuring symptom validity during an evaluation, the *WMT* task differs because it relies more upon verbal memory whereas the *TOMM* task generally relies upon visual memory. The *WMT* has been advertised as a test that is almost solely sensitive to poor effort. Flaro, Green, and Robertson (2007) have noted that the symptom validity components of the *WMT* were designed to be "virtually insensitive to all but the most extreme forms of impairment of learning and memory" (p. 374).A number of studies have been conducted that investigate how sensitive the *WMT* is to anticipated factors such as poor effort or unintended factors such as genuine cognitive dysfunction, reading level, or age. In one study, Green and Flaro (2003) make the case that the *WMT* measures symptom validity rather than ability by demonstrating that children from age 7 to age 18 perform similarly on this test as long as they possess a third grade reading level. However, Allen, Bigler, Larsen, Goodrich-Hunsaker, and Hopkins (2007) found via fMRI that the *WMT* activates dorso-lateral prefrontal cortex, superior parietal lobes, and the anterior cingulate and concluded that the *WMT* requires "significant" cognitive effort. Larsen, Allen, Bigler, Goodrich-Hunsaker, and Hopkins (2010) replicated the study of small group of young adult men ($n = 4$) by Allen et al. (2007) using a larger sample size ($n = 10$) which included an equal number of young adult men and women. Larsen and colleagues (2010) found very similar patterns of brain activation in the replication study during the delayed recognition portion of the *WMT* and did not find sex differences. In a study with three amnestic patients with known hippocampal brain damage and four matched controls, Goodrich-Hunsaker and Hopkins (2009) found that performance on each task of the *WMT* was approximately three standard deviations lower than the mean performance of adult control participants. Interestingly, although the amnestic patients were clearly impaired with their *WMT* performance, they all still scored above the *WMT* cut scores and therefore were categorized as giving adequate effort.

The literature suggests that symptom validity tests can be sensitive to the effects of substances. In a 2011 study, Loring, Marino, Drane, Parfitt, Finney, and Meador conducted a double-blind,

placebo-controlled, crossover study which investigated how a 2 milligram dose of lorazepam impacted performance on the *WMT*. Loring and colleagues (2011) found that 6 of the 28 participants in the study failed the *WMT* while on lorazepam whereas they were able to pass the *WMT* under a placebo condition; 1 of the 28 participants failed the *WMT* under placebo condition but not while taking lorazepam in the experimental condition. It was concluded that variables such as lorazepam can impact functioning on cognitive tests and on symptom validity tests and that this should be taken into account while clinicians may decisions (Loring et al., 2011).

Some symptom validity tests may be more vulnerable to distraction. Batt, Shores, and Chekaluk (2008) found that the *TOMM* was more resilient to distraction compared to the *WMT* and therefore may be less impacted by some forms of cognitive disability. Batt and colleagues (2008) designed a study utilizing non-litigating participants with either traumatic or non-traumatic brain injury. The participants were randomly assigned to Simulated Malingering (n = 11), Distraction (n = 24), or Full Effort (n = 25) groups and were given the *TOMM* and *WMT* in counterbalanced order. The individuals instructed to give full effort were given standard instructions and a standard administration of the *TOMM* and *WMT*. The individuals in the distraction condition were asked to orally add 3 to a number presented every 3 seconds throughout the duration of the learning portions of the *WMT* and *TOMM*. The individuals in the simulated malingering group were given instructions to fake a memory impairment without faking to a degree that it was obvious. Batt and colleagues (2008) found that *WMT* performance was significantly impacted by distraction whereas *TOMM* performance was not. Interestingly, both tests were 100% sensitive to the simulated malingering; however, specificity for the *TOMM* under a full effort condition was 84% and for the *WMT* it was 56%. The literature reveals that there is some argument over the superiority of symptom validity tests such as the *TOMM* and *WMT*. Martins and Martins (2010) described the *WMT* as one of the best tests for memory malingering despite some noted concerned about its false positive rate. While some researchers have expressed concern

about the false positive rate of the *WMT*, other researchers have found ways to explore and address these concerns. Green, Montijo, and Brockhaus (2011), while considering a "dementia profile" (p.92) for people with probable dementia concluded the false positive rate of the *WMT* was no higher than 1.6%. Green and colleagues (2011) formed a study with 60 participants who had probable, mild, moderate dementia or no impairment based upon a *Clinical Dementia Rating* scale (*CDR*; Morris, 1993). Participants were given the *WMT* with the option of having the examiner read stimulus items and control the computer mouse. The authors noted that the interval between trials was also shortened to reduce task demands. The examiners found that all groups of participants had differing levels of performance on the *WMT* depending upon level of impairment. It was also found that scores on a simpler task, the *MSVT*, were higher for all groups of participants. Using profile analysis, Green and colleagues (2011) were able to eliminate all cases of false positives in the dementia groups in the study.

Greiffenstein, Greve, Bianchini, and Baker (2008) wrote that some of the research concluding the superiority of the *TOMM* is erroneous due to methodological flaws. Green (2007) also suggested the *WMT* has properties which make it superior to the *TOMM*. The agreement rate between the *TOMM* and *WMT* is 77.2% according to Greiffenstein and colleagues (2008). Greiffenstein and colleagues (2008) tested the agreement rate by using data from 473 people who were in neuropsychological evaluations for the purposes of receiving compensation due to possible cognitive or persistent pain disability related to neurological trauma. The *TOMM* and *WMT* were administered in each of the neuropsychological evaluations. The administration and interpretation of the *TOMM* and *WMT* were standard aside from the cut scores from Trial 2 of the *TOMM* being applied to all three parts of the *TOMM* to make *WMT* and *TOMM* comparisons symmetrical. The authors found that in 13.7% of cases, examinees failed the *WMT* and passed the *TOMM*; in 9.1% of cases, examinees failed the *TOMM* but passed the *WMT*. Greiffenstein and colleagues (2008) concluded that especially after observing cut scores on all three parts of the *TOMM*, there is no reason to conclude that either the

TOMM or *WMT* is superior. The agreement rate is high enough on tests like the *TOMM* and *WMT* to demonstrate the usefulness of symptom validity tests, but low enough to demonstrate the need for neuropsychologists to consider these tests in the context of other available information. Flaro, Green, and Robertson (2007), using participants with incentives to perform well on testing, incentives to appear impaired on testing, and no significant incentives to modify performance found the expected patterns of performance on the *WMT* with those having incentive to appear impaired failing the *WMT* more frequently.

Bauer, O'Bryant, Lynch, McCaffrey, and Fisher (2007) evaluated the efficacy of shortened versions of the *WMT* and the *TOMM*. Bauer and colleagues (2007) utilized archival data from 64 examinees who were litigants with mild traumatic brain injuries. The authors administered the *TOMM* and *WMT* according to standard instructions and used the classification based upon the *TOMM* and *WMT* manuals as the criteria for assessing accuracy of shortened versions of the tests. For both the *WMT* and for the *TOMM*, using just the *WMT*'s Immediate Recall or just the *TOMM*'s Trial 1 resulted in high sensitivity and specificity. Bauer and colleagues (2007) concluded that using the shortened versions of the *TOMM* or *WMT* could result in useful screening measures as sensitivity and specificity over 0.8 were possible for either measure if correct cut scores were used.

Test security is a concern for symptom validity tests as well as some of the other tests developed for neuropsychological assessment. Examinees who have been able to study test materials or receive coaching prior to an evaluation may be able modify their performance in such a way that their results are invalid. Unfortunately, it appears that information on symptom validity tests may be readily accessible to some examinees. Bauer and McCaffrey (2006) found that enough information on the *TOMM*, *VSVT*, and *WMT* was located on the Internet to potentially threaten the validity of the tests. With the help of a graduate student, Bauer and McCaffrey (2006) used the search engine "Google" on two different occasions to examine the top 50 results on the three symptom validity tests. The authors

found that over the course of four months, some of the search results had changed; the more recent results were used for analysis. Depending on specificity of test descriptions on the websites, four different threat levels were assigned. Although Bauer and McCaffrey (2006) concluded that the *TOMM* search results had the highest number of "high" threat level websites, they gave examples of some specific alarming findings such as a readily accessible sample psychological report with *VSVT* score ranges and classifications. On the *TOMM*, *VSVT*, and *WMT*, various information on cut-off scores, recording of reaction times, and item difficulty was readily accessible on the Internet (Bauer & McCaffrey, 2006).

The possibility of disclosure of sensitive test information may in some cases be unintentional. Psychologists, attorneys, and laypeople may have some access or familiarity with test information and be unaware of the problems associated with disclosing this information. Morel (2009) posits that applied neuropsychology has not issued guidelines to attorneys regarding test information that can and cannot be shared with clients and therefore the attorneys sometimes disclose information without knowledge of how it can impact the validity of neuropsychological examinations. It is likely that many legal professionals would be receptive to receiving information on test security of psychological measures as evidenced by statements made by U.S. Supreme Court justices since at least the late 1970's (Morel, 2009). Morel (2009) provided comprehensive guidelines for how attorneys can appropriately prepare a client without threatening validity of neuropsychological evaluation results.

Prior to neuropsychological evaluation, there is a potential that examinees may be warned of the presence of symptom validity indicators or even coached on how to perform on certain tests. Although coaching may cause examinees to modify their behaviors and responses during a neuropsychological evaluation, some symptom validity tests have been shown to be resilient to coaching. Davis, Wall, & Whitney (2012) noted differences in performance on the *WMT* and *TOMM* for simulated naïve and simulated coached malingering groups; a second clinical validation study using data from actual VA

hospital neuropsychological examinations was also conducted. Davis and colleagues (2012) developed three consistency scales for the *TOMM* that measured if examinees were consistently correct or incorrect with responses across Trial 1, Trial 2, and the Retention Trial. Essentially, it was found that malingerers with more coaching appear to be able to demonstrate a performance that is impaired yet more believable. Coached malingerers may have better performance on symptom validity tests compared to naïve malingerers, but on consistency indices of measures such as the *TOMM* or *WMT*, coaching does not appear to improve performance (Davis et al., 2012). The findings of Davis and colleagues (2012) suggest that examining consistency of responses on the *TOMM* may help detect malingering in cases where examinees are coached. Jelicic and colleagues (201), using a sample of participants simulating malingering (n = 90) found that all participants instructed to give full effort were correctly classified, 97% of symptom-coached participants were correctly identified, and 87% of participants who were symptom and test-coached were correctly identified. The findings of Jelicic and colleagues were consistent with previous studies such as a study by Powell, Gfeller, Hendricks, and Sharland (2004). Although findings from studies with simulators may not perfectly generalize to clinical populations, the current evidence suggests that symptom validity tests, such as the *TOMM* will often still be efficacious despite coaching.

Dedicated symptom validity test performance is often related to overall performance on neuropsychological tests. Measured effort explained 53% of the variance in a study examining 30,736 individual test scores (Green, Rohling, Lees-Haley, & Allen III., 2001). In a study of 904 patients, including 470 patients with head injuries, Green and colleagues (2001) analyzed data from an average of 34 neuropsychological tests plus 2 symptom validity indicators. Green and colleagues (2001) created an overall Symptom Validity score from scores on the *Computerized Assessment of Response Bias* (*CARB*; Allen, Conder, Green, & Cox, 1997), the *Word Memory Test* (*WMT*; Green, 2003), and the *California Verbal Learning Test* (*CVLT*; Delis, Kramer, Kaplan, & Ober, 1987) and compared this

validity score to an Overall Test Battery Mean which was designed to give an overall score to represent the global functioning of the individual by combining the average score of the patient across all measures. It was found that the correlation between the Symptom Validity score and the Overall Test Battery Mean was 0.74 (Pearson's *r*). Green and colleagues (2001) found that the Symptom Validity score was a better correlate of overall performance compared to any other measured domain. This study strongly suggests failure of symptom validity tests will likely indicate that scores on other neuropsychological tests are invalid.

Gunner and colleagues (2012) found that suboptimal performance on symptom validity tests are related to decreased performance on neuropsychological memory tests. Gunner and colleagues (2012) analyzed data from 48 patients who had evaluations for medico-legal reasons. It was noted that all but 5 patients had mild traumatic brain injury histories. Gunner and colleagues used data from the *HRNB* (Reitan and Wolfson, 1993) the *Memory Assessment Scales* (*MAS*; Williams, 1991), the *TOMM* (Tombaugh, 1996), and the *WMT* (Green, 2003). Gunner and colleagues (2012) used symptom validity test data to make a determination of optimal and suboptimal effort based upon *TOMM* performance alone, based upon *WMT* performance alone, based upon data from both the *TOMM* and *WMT*, and an alternate interpretation of the *TOMM* known as the Albany Consistency Index (ACI; Gunner et al., 2012). It was found that 14.5% of the variance on the *MAS* Global Memory Index, which is a summary index of Short-term Memory, Verbal Memory, and Visual Memory, could be accounted for by the ACI and 26.6% of the variance of the *MAS* Global Memory Index could be accounted for using just *WMT* criteria. The relationship between the Global Memory Index and standard *TOMM* scores was nonsignificant. Although variance accounted for Green and colleagues (2001) study and the Gunner and colleagues (2012) study had a noteworthy difference, it appears that the studies agree that performance on symptom validity tests are related to performance on neuropsychological tests.

There is data suggesting that performance on symptom validity tests can relate to performance on measures of personality. Whitney (2012,), for example, found that the Response Bias Scale on the *Minnesota Multiphasic Personality Inventory-2* (*MMPI-2*; Butcher et al., 2001) accounted for about 20% of the variance on *TOMM* scores. This underscores that symptom validity test performance can be related to performance on almost all parts of a neuropsychological evaluation.

Structured Interviews and Inventories Sensitive to Malingering

Self-report behavior and personality rating scales can also be useful for detecting malingering. The *Minnesota Multiphasic Personality Inventory – 2* (*MMPI-2*) can serve as a useful tool for determining if an examinee is falsely representing pathology (Toomey et al., 2009). Scales on the *MMPI-2* should not be considered in isolation as validity scales on the *MMPI-2* accounted for less than 26% of the variance on a *Medical Symptom Validity Test* (*MSVT*; Green, 2004) and less than 20% of the variance on the *TOMM* (Tombaugh, 1996) according to Whitney (2013). Novo and colleagues (2013) established the *MMPI-A* had an excellent ability to classify adolescents who were malingering.

The Millon Clinical Multiaxial Inventory – III (*MCMI-III*; Millon, 1994) has demonstrated efficacy with detecting malingers; each validity indicator was able to detect about 50% of malingers in a group of TBI patients (Aguerrevere, Greve, Bianchini, & Ord, 2011). Aguerrevere and colleagues (2011) utilized as participants 108 referrals to a psychology clinic. The 108 referrals had varying degrees of reported TBIs from mild to severe and all but six referrals had a significant external incentive. One participant was excluded due to psychiatric conditions. The participants were administered the MCMI-III in a standard manner and were classified as having malingered neurocognitive dysfunction (MND), not having MND, or being in an indeterminate group using criteria developed by Slick and colleagues (1999). To qualify as having MND, by the criteria of Slick and colleagues (1999), individuals must have an external incentive present (Criteria A), evidence from neuropsychological testing (Criteria B), evidence from self-report (Criteria C), and the presence of

behaviors in Criteria B and Criteria C must not be fully accounted for neurological, psychiatric, or developmental factors (Aguerrevere et al., 2011). Aguerrevere and colleagues (2011) found that all of the *MCMI-III* modifier indices were useful in detecting malingerers and that by using all indices, 56% of malingerers were detected with no false positives.

Structured interviews can be effective in identifying malingering. According to Toomey and colleagues (2009), despite imperfect classification accuracy, the *Structured Interview of Reported Symptoms* (*SIRS*; Rogers, Bagby & Dickens, 1992) is regarded by some to be the most accurate measure for detecting malingering. Unfortunately, structured interviews like the *SIRS* and open-ended interviews can sometimes take considerable time and resources to complete; therefore it is in the interest of clinicians and patients to use the most efficient and appropriate set of measures for the circumstances. Smith and Burger (1997) suggested using malingering screening measures like the *Structured Inventory of Malingered Symptomatology* (*SIMS*) and examining results of personality inventories to screen for malingering prior to administration of extended measures such as the *SIRS*. Smith and Burger (1997) reached this conclusion via a study with college students (n = 476) divided into eight groups with various conditions of simulated malingering and full effort responding. It was found that the *SIMS* had the highest level of sensitivity when compared to the F and K scales of the *Minnesota Multiphasic Personality Inventory* (*MMPI*), the 16PF Faking Bad scale, and portions of the malingering scale (Smith & Burger, 1997). Smith and Burger (1997) pointed out the limitations of the study which included the fact that college students were asked to simulate malingering for some extra credit.

Conclusions

Malingering, which is often generally defined as exaggerating or feigning impairment in order to gain an external incentive, is a condition that neuropsychologists must be mindful of during the evaluation process. Malingering has been present in some form throughout most of human history with

some of its earliest roots being in military history. Malingering can be relatively uncommon in general healthcare settings but rather routine in forensic settings or other settings where notable incentives to malinger exist. Malingering detection was often done by clinical instinct until more literature about using more objective tests appeared in the late 1970's and beyond. It is clear that malingering can impact the results of many common assessments used in neuropsychological evaluations, yet the literature suggests that analyzing test scores in the absence of data specifically related to symptom validity is often insufficient for detection of malingerers.

A number of stand-alone symptom validity tests and embedded validity indicators exist to help a practitioner identify malingering; however, experts in the field routinely encourage practitioners to use the data in context of an entire evaluation. Observations of the examinee, tests of physical functioning, and careful examination of medical records can sometimes aid in the detection of malingering. Identification of malingerers can be difficult due to the imperfect measures available, examinees exhibiting malingering inconsistently during the evaluation, examinees being coached on how to defeat symptom validity measures, examiner mindsets, and examinees giving altered effort unintentionally.

Research and practice related to malingering continues in hope of identifying better ways to classify malingering. Currently, the literature provides a number of studies validating longstanding measures of symptom validity and some comparisons amongst measures, but to date the relationship between performance on dedicated symptom validity tests and some commonly administered neuropsychological tests has not been thoroughly elucidated.

Chapter III

Methodology

The purpose of this chapter is to detail how participants were selected and describe procedures that were used to gather and analyze data. This chapter is divided into four parts: Participant Selection; Data collection procedures; Instrumentation, Validity, and Reliability; and Data Analysis.

Participant Selection. Participants for the study were recruited through a pool of general psychology students seeking to participate in psychological research; the research pool from Psychological Sciences at Ball State University was used. Participants consisted of university students over the age of 18 who are enrolled in a psychology course at the undergraduate level. A total of 46 participants were recruited for the study.

Procedures. All data collection from participants occurred in a manner consistent with procedures approved the university's Institutional Review Board (IRB) for a study for which Dr. Andrew Davis is the primary investigator; Evaluating the Relationship between Memory, Intellectual Functioning, Executive Functioning, Language, Visual-Spatial/Construction, Attention and Effort. The writer of this study and other graduate students participated in data collection using procedures approved by the IRB. Data collection occurred in quiet, well-lit rooms. Upon arrival participant arrival at a room, examiners reviewed the informed consent form with the participants who then signed an informed consent form. Participants provided demographic data. Collected demographic

information included participants' age at testing, gender, ethnicity, handedness, level of education, parents' level of education, parents' occupation, current medications, psychiatric diagnoses, learning disabilities, and history of brain injury. Prior to test administration, participants were read standard statements informed them of the nature of the following tasks and encouraged participants to put forth full effort during their session. Through the reading of standard instructions, participants were informed that they may take breaks whenever they wish and that they could withdraw from the study at any time and for any reason.

Participants were administered neuropsychological assessments in a manner consistent with the standardized instructions present in the respective examiner's manual of each assessment. Participants were individually administered assessments in a quiet, well-lit room. All participants were administered assessments by advanced graduate students who had successfully completed a class in cognitive assessment and additional training for administration of the specific assessments used in the study. The current study examined results from a subset of the instruments in Dr. Davis' study, the *Test of Memory Malingering* (*TOMM;* Tombaugh, 1996), *Word Memory Test* (*WMT;* Green, 2003), and the *Wechsler Memory Scale - Fourth Edition* (*WMS-IV;* Wechsler, 2009). Administration time for the *WMS-IV* (Wechsler, 2009) is generally 90 minutes for the primary subtests (Drozdick, Holdnack, & Hilsabeck, 2011). The administration time of the *WMT* (Green, 2003) is generally 10 to 15 minutes. The administration time of the *TOMM* (Tombaugh, 1996) is generally 15 to 20 minutes. The administration time for the above measures factored into an expected total administration time of approximately four hours for the larger study. Participants received 4 hours of research participation credit for their psychology course for taking part in the study.

Instrumentation

Test of Memory Malingering. The *Test of Memory Malingering* (*TOMM*; Tombaugh, 1996) is an assessment employed by psychologists to help determine if examinees are suffering from memory

impairment or providing insufficient effort. The test appears to be a visual memory recognition task with two learning trials. The test is designed to be resistant to a wide range of neurological impairments and sensitive to insufficient effort. The *TOMM* can be administered by hand or via computer format. The *TOMM* is administered to individual examinees and can be administered to individuals from 16 to 84 years of age. The administration time for the *TOMM* is 15 to 20 minutes for standard administration.

The development of the *TOMM* was based upon research in the fields of cognitive psychology and neuropsychology. In order to detect malingering behavior, the test was designed to appear difficult despite actually being simple. The development of the *TOMM* was based upon research, such as that conducted by Standing, Conezio, and Haber (1970), that demonstrated individuals (n = 2) who were shown 1,100 pictures for 5 seconds each could each recognize over 95% of presented pictures after a 30 minute delay when these pictures were presented in pairs consisting of new and previously viewed pictures.

According to Tombaugh (1996) the initial normative testing for the *TOMM* was conducted on 405 individuals who ranged in age from 16 to 84 years with an average of 13.1 years of education. This non-clinical sample was recruited via word of mouth and in public places such as shopping centers. Some participants received university course credit although none of the participants received financial compensation. Although the *TOMM* was modified after the initial normative testing, the sample could identify 94% of target images on the first trial and over 99% of target images on the second trial and retention trial. About 7.5% of variance on Trial 1 was accounted for by education and less than 2% of variance on the second trial and the retention trial was accounted for by education. On all *TOMM* trials, individuals performed at levels higher than expected.

After adjustments were made to the *TOMM*, such as changing the number of distractors and changing the level of feedback, more normative data was collected from a non-clinical sample of 70

individuals with an age range of 17 to 73 years. Participants in this sample had an average of 12.7 years of education. Comparable to the first phase of normative testing, participants had 95.6% accuracy on Trial 1 and over 99% accuracy on the other two trials. Both phases of normative data collection demonstrated that *TOMM* performance for individuals without neurological impairment is high.

Validation of the *TOMM* with a clinical sample was conducted using scores from 158 patients (after excluding data from 3 patients who were too impaired to test). This clinical sample consisted of individuals with no cognitive impairment (n = 13), cognitive impairment (n = 42), aphasia (n = 21), traumatic brain injury (n = 45), and dementia (n = 40). No patients in the clinical sample were involved in litigation or compensation hearings. The resistance of the *TOMM* to influence from bona fide impairment was demonstrated by the findings of no significant difference in performance among the first four listed groups using an ANOVA and Tukey-HSD posttest. The same analysis revealed the dementia group was significantly different; however, the dementia group was still able to obtain higher than 92% accuracy on Trial 2 compared to over 97% accuracy obtained by all of the other groups.

To explore characteristics of the *TOMM* while being used with potential malingerers, a sample of 49 undergraduate students without known neurological impairment were recruited. Students were either given instructions to perform with full effort (n = 22) or asked to simulate malingering (n = 27). The students asked to simulate malingering were given a week to prepare and were given a detailed scenario which provided coaching on how to convincingly demonstrate impairment. Students were offered a reward of $50 for successfully malingering. As expected the group instructed to perform with full effort generally had extremely high performance on the *TOMM* consist with the non-clinical normative studies described previously. Using an accuracy cut-off score, 82% of simulated malingerers were correctly identified and 100% of students in the control group were correctly identified as providing adequate effort (Tombaugh, 1996).

Another study was conducted on a clinical sample with some patients who were "at-risk" for malingering. This sample consisted of TBI (Traumatic Brain Injury) patients who were not "at-risk" for malingering (*n* = 17), TBI patients "at-risk" for malingering (*n* = 11), cognitively intact controls (*n* = 11), and patient controls with significant focal neuropsychological impairment recruited from a neurological unit (*n* = 12). On Trial 2 and the Recognition trial, analysis with ANOVA and a Tukey-HSD posttest revealed no significant difference among the three groups not determined to be "at-risk" for malingering. The group "at-risk" for malingering, based upon meeting one or more predetermined criteria, had significantly different performance from the other three groups. It was also determined that performance on select parts of the *California Verbal Learning Test and Wechsler Memory Scale – Revised* were relatively independent from *TOMM* performance which suggested little overlap in the constructs being measured.

Word Memory Test. The *Word Memory Test* (*WMT*; Green, 2005) is an assessment utilized by psychologists to measure verbal and nonverbal memory. The *WMT* consists of multiple subtests intended to measure memory; these subtests contain embedded measures intended to determine validity of examinee performance. The test is designed to be resistant to a wide range of neurological impairments and sensitive to insufficient effort. The *WMT* is administered via computer; however, an oral form of the *WMT* is available for special circumstances. The *WMT* is administered to individual examinees and can be administered to adults and children who have at least a 3rd grade reading level. The administration time for the *WMT* is 10 to 15 minutes.

The *WMT* normative group was collected by several practitioners in independent clinical settings in the U.S. and Canada. The main normative sample consisted of data from 1,250 patients with some patients reporting head injury (*n* = 535), neurological patients with strokes, aneurysms, multiple sclerosis, and other conditions (*n* = 89), patients with major depression (*n* = 85), anxiety disorders (*n* =

18), orthopedic injuries (*n* = 77), chronic fatigue syndrome (*n* = 34), chronic pain syndrome or fibromyalgia (*n* = 61), and other conditions (*n* = 101).

Validation of the *WMT* was based upon a study (Iverson, Green, & Gervais, 1999) with 40 healthy volunteers and 57 patients with moderate to severe brain injuries. Healthy volunteers had 97.8% accuracy on the three *WMT* symptom validity measures whereas patients with brain injuries had 95.1% accuracy. This demonstrated relative insensitivity of the *WMT* to known impairment.

Another study conducted by Green & Allen (1999) with 40 neurological patients divided into high impairment and low impairment groups based upon *California Verbal Learning Test* impairment (*CVLT;* Delis, Kramar, Kaplan, & Ober, 1987) found no significant difference in *WMT* performance. Green (2005) noted that *WMT* subtests designed to measure memory were significantly lower in patients with impaired *CVLT* scores compared to patients with normal range *CVLT* scores.

Gorissen, Sanz de la Torre, and Schmand (2003) found that German and Spanish translations of the *WMT* provided results consistent with the neurological patients and controls tested with the English version of the *WMT*. It was found that for all three symptom validity measures, the Immediate Recall (IR), Delayed Recall (DR), and Consistency Scores, the control and neurologically impaired groups did not differ significantly in performance.

Green, Iverson, and Allen (1999) found that in a large sample, individuals with greater head injury severity scores perform at a higher level than individuals with milder head injuries. This suggested that the *WMT* was resistant to the effects of impairment and sensitive to exaggeration of deficits.

Green (2005) noted that a study of 20 healthy volunteers provided test-retest reliability for the subtests of the *WMT* at $r = 0.92$ to $r = 0.99$ based upon subtest; however, in a clinical sample of 33 patients tested with the *WMT* two times over the course of a year or more the test-retest reliability for IR and DR was 0.43 and 0.33 respectively. It was suggested that the modest correlation was related to

varying effort rather than unreliability. Cases with effort being satisfactory during testing and unsatisfactory during retesting were noted during the study of test-retest reliability of the clinical sample (Green, 2005).

Iverson, Green, and Gervais (1999) demonstrated efficacy of the *WMT* with a simulator study. In this study, 25 highly educated and unimpaired individuals were asked to take the *WMT* and simulate genuine memory impairment without exaggerating to a point of being detectable. While 40 healthy volunteers instructed to give full effort had an average of 97.8% accuracy on the *WMT*, the 25 simulators had an average of 67.2% accuracy on the *WMT* symptom validity tests. The simulators indicated trying to be subtle with impairment but scored more than 5 standard deviations lower than neurological patients on the DR measure.

Green (2005) also cited another study with 20 patients who passed the *WMT* during a disability examination with 98.2% correct on average for the DR measure. The individuals were asked to retake the *WMT* and simulate memory impairment with care being taken to not exaggerate to a degree that would be detectable. The average performance upon retest was 62.6% accuracy.

Green (2005) noted that the sensitivity of the *WMT* to exaggeration was 100% during the clinical study and 96% on the simulator study with well-educated volunteers. All individuals instructed to give full effort on this simulator study were correctly classified as providing adequate effort. The classification cutoff scores were set much lower than average scores of samples with brain injuries and samples of children with psychiatric illness so that the expected false positive rate for the *WMT* is expected to be extremely low.

Wechsler Memory Scale – Fourth Edition. The *Wechsler Memory Scale – Fourth Edition* (*WMS-IV*; Wechsler, 2009) is an assessment employed by psychologists to comprehensively test memory of individuals who have been diagnosed or suspected to have any number of neurological or psychiatric conditions. The *WMS-IV* has subtests that can be added or removed from an assessment to

compose relevant test batteries. The *WMS-IV* is administered orally to individual examinees and has normative data available for individuals who are 16 to 90 years old. The administration time for the *WMS-IV* is 90 minutes for the primary subtests (Drozdick, Holdnack, & Hilsabeck, 2011). According to the technical and interpretative manual of the *WMS*-IV (Wechsler, 2009), the normative sample for the *WMS-IV* normative sample consisted of 1,400 individuals. There were 100 individuals divided in 14 age bands from age 16 to 90. Adults in the older age bands were administered a briefer battery designed for older adults. Individuals in the normative sample were picked in such a way that individuals in each age band represented demographic data for gender and ethnicity on the 2005 U.S. Census. The normative sample was also stratified based upon education level. The sample was obtained from 480 examiners representing all of the major geographic regions of the U.S. Start points and discontinue rules were set to ensure that examinees would not be exposed to a large number of excessively easy or excessively hard items. Start points were set at a level where 95% of people at a given age would pass the initially presented items.

Reliability of the *WMS-IV* was computed using split-half reliability based upon the Spearman-Brown formula (Crocker & Algina, 1986; Li, Rosenthal, & Rubin, 1996) and internal consistency reliability for the *WMS-IV* was calculated using a formula from Nunnally and Bernstein (1994). Average reliability coefficients were calculated using Fisher's z transformation (Silver & Dunlap, 1987; Strube, 1988). Internal consistency was not calculated for Verbal Paired Associates Word Recall Scaled Score and recognition memory measures as the number of items available to evaluate vary from individual to individual.

The average split-half reliability for the *WMS-IV* index scores on the Adult battery ranged from 0.93 to 0.96; similar reliability was reported for a sample of 555 individuals with a variety of clinical disorders (Drozdick, Holdnack, & Hilsabeck, 2011). Reliability for subtests ranged from 0.82 to 0.97

in the Adult battery with slightly higher values found for the clinical sample (Drozdick, Holdnack, & Hilsabeck, 2011).

Recognition memory tasks on the *WMS-IV* had restricted score ranges, so decision-consistency reliability was utilized to demonstrate reliability.

WMS-IV Verbal Paired Associates I, Visual Reproduction I and II, and Spatial Addition had the highest internal consistency or reliability with other subtests being in the moderate to high range. An examination of 555 individuals diagnosed with at least one of a number of psychiatric or neurological impairments revealed a high level of internal consistency for the *WMS-IV* and suggested its generalizability.

The test-retest stability was examined by giving 244 individuals the *WMS-IV* on two occasions. The test-retest interval was from 14 to 84 days with a mean of 23 days. The demographics of the sample were comparable to the general population demographics. The corrected r for the first and second testing ranged from .59 to .81 based upon subtest. The correction to r was applied to account for variability in the normative sample. Some of the score differences in test-retest reliability were assumed to have occurred due to practice effects. Interscorer agreement on more objectively scored parts of the *WMS-IV* ranged from .98 to .99. On subtests with more subjective scoring guidelines (e.g. Clock Drawing), agreement was at 96% between two independent scorers.

Validity was in part established by correlating the *WMS-IV* with the *WMS-III*. A study examined the performance of individuals age 17 to 69 on both measures. A total of 224 examinees were given both tests between 13 and 98 apart. Since parts of the *WMS-IV* were significantly reworked, some indices had corrected correlation coefficients of .50 whereas some correlated at the .81 level. As expected, correlation of subtests with comparable subtests of the other measure was as low as .35. Correlations with other popular memory tests, intelligence tests, and measures of interest were also computed to examine the concurrent validity of the *WMS-IV*. Flanagan and Harrison (2012) note

that the correlations between the *WMS-IV* and *WMS-III* are high, but the reworking of visual components of the test lower correlations of the visual tests. Correlation between the *WMS-IV* and measures examining similar constructs, such as the *Children's Memory Scale* (Cohen, 1997) are high, whereas correlations between the *WMS-IV* and intelligence, achievement, or other measures are low to moderate (Flanagan & Harrison, 2012). As an example, the *WMS-IV* and a measure of achievement, the *Wechsler Individual Achievement Test* (*WIAT-II*; Wechsler, 2005) have correlation coefficients ranging from 0.29 to 0.77 based upon indices measured. Construct validity was assessed via examining intercorrelations of subtest scores and comparing solutions of factor-analytic studies. The technical and interpretive manual of the *WMS-IV* provides strong evidence for construct validity of the Auditory Memory Index (AMI), Visual Memory Index (VMI), and Visual Working Memory Index (VWMI); however, studies, such as a study conducted by Hoelzle, Nelson, and Smith (2011) found just two factors via principal component analysis conducted using data from the *WMS-IV* normative sample. Hoelzle and colleagues (2011) found that Logical Memory I & II and Verbal Paired Associates I & II load on an auditory component at 0.62 or higher whereas Designs I & II, Visual Reproduction I & II, Spatial Addition, and Symbol Span load on a visual component at 0.61 or higher; no subtest loaded on the other factor at higher than 0.24. Content validity was assessed via direct questioning examinees on their response processes and looking for items that was often answered by respondents in an incorrect manner. Adjustments to the test instructions, distractors, and items were made during test construction as needed.

Restatement of Research Questions

R_1 How much variance in the WMS-IV subtests does the TOMM account for?

R_2 How much variance in the WMS-IV subtests does the WMT account for?

R_3 How much variance in the WMS-IV composites does the TOMM account for?

R_4 How much variance in the WMS-IV composites does the WMT account for?

R_5 Does the examination of the means of all variables in the obtained sample look comparable to the means found in the literature?

R_6 How much correlation is present between TOMM and WMT scores?

R_7 How much correlation is present between WMS-IV, TOMM, and WMT scores?

Data Analysis

To understand the demographics and performance characteristics of the sample, descriptive statistics were used to understand average performance and the standard deviation for performance on the WMS-IV and subtests of the TOMM and WMT.

Computation of a series of bivariate Pearson *r* coefficients were used to understand how each TOMM and each WMT subtest relate to the auditory memory and visual memory subtests of the WMS-IV in a univariate fashion, prior to examining these relationships in the context of the full model to be described below.

To address research questions, a type of Structural Equation Modeling (SEM), known as path analysis was planned. SEM is a set of statistical tools that allows for examination of how independent and dependent variables relate (Ullman, 2013). The adequacy of SEM models for research questions was examined and presented using IBM® SPSS® Amos 19.0. Models were presented to address the research questions to reduce model complexity and aid in the interpretation of how both autonomous symptom validity measures, the TOMM and WMT predicted performance on the WMS-IV. As indicated in the figures, it was hypothesized that performance on the TOMM and WMT impacted performance on the WMS-IV rather than WMS-IV performance impacting the TOMM and WMT performance. This directionality was assumed due to studies indicating that performance on the TOMM and WMT are generally not impacted by an array of conditions related to mental disorders or head injuries. While usage of IBM® SPSS® Amos 19.0 was originally planned, relationships between

measures was explored using R i386 3.3.2 (R Core Team, 2013) with relevant add-on packages due to resource availability.

Due to the TOMM and WMT consisting of tasks that appear to be assessing visual and auditory memory, respectively, it was expected that the relationships between each symptom validity measure and the subtests of the WMS-IV would be different. It was expected that the TOMM would account for more variance in the visual memory subtests of the WMS-IV whereas the TOMM would account for more variance in the auditory memory subtests of the WMS-IV.

In the models, as depicted in the figures, all variables were treated as observed variables. The relationship between measures was explored primarily through multivariate regression. Latent variables such as effort or the various types of memory were assumed to be represented by subtest performance and therefore were not be included in the models. Multivariate regression often works best with sample sizes in the hundreds or thousands. Techniques developed for statistics with low sample size (e.g. Bentler & Yuan, 1999) were considered while working with data acquired from the 46 participants in the study.

In order to address the first four research questions, multivariate regressions were used. See figures 1 through 20. The multivariate regressions explored how each predictor variable from the TOMM and WMT related to each observed variable of the WMS-IV subtests and WMS-IV composites.

In order to address research question five, the means and standard deviations were computed for each predictor variable and each observed variable for the TOMM, WMT, and WMS-IV. These values were explored in relationship to values obtained in the examiner's manuals of each measure and other relevant literature available.

Research questions six and seven were explored by using Pearson and Spearman correlations. These values were compared to the literature as well as data obtained through the multivariate regressions used in the previous questions.

Figure 1: Regression model exploring how TOMM Trial 1 relates to WMS-IV subtests.

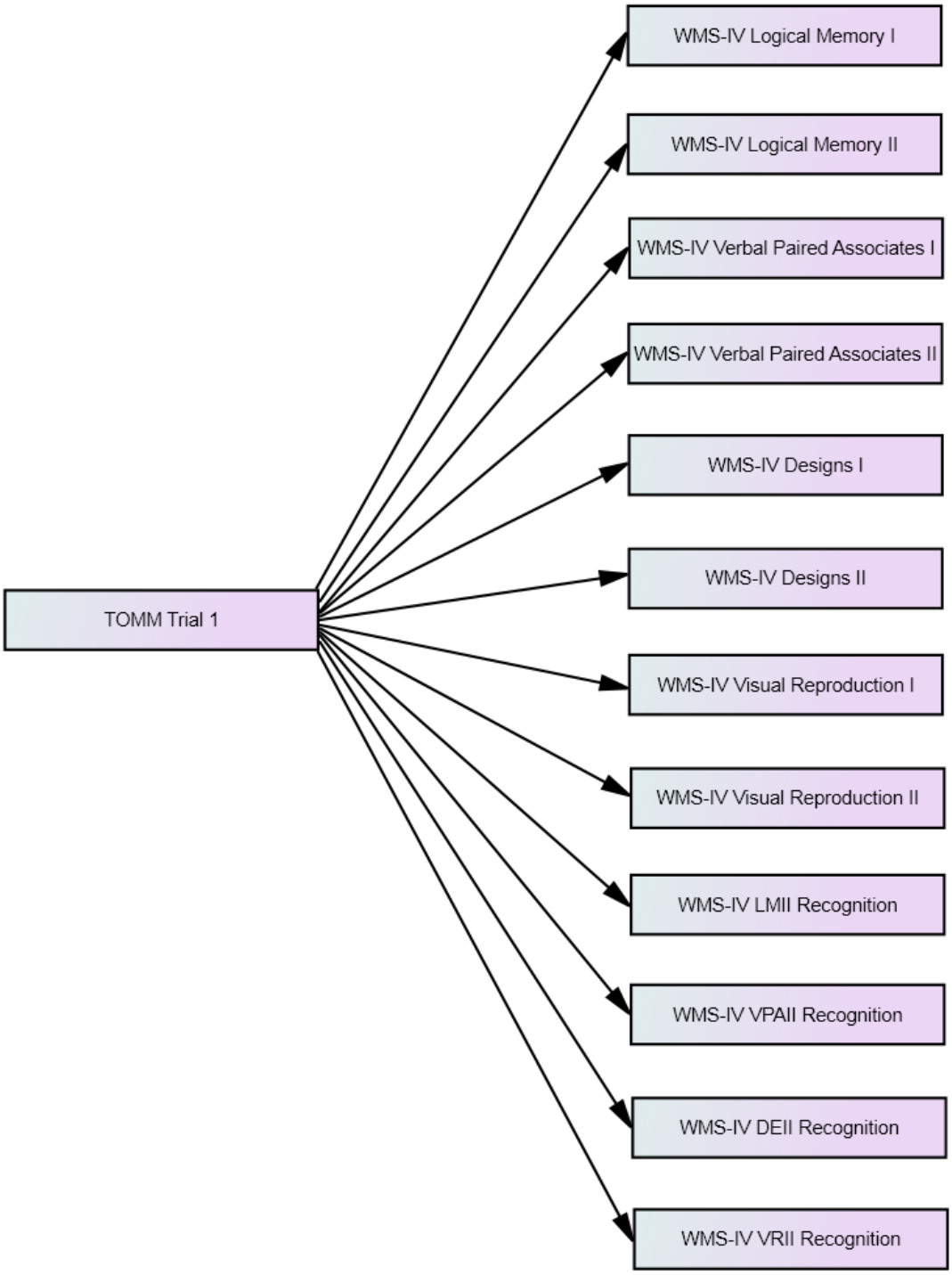


Figure 2: Regression model exploring how TOMM Trial 2 relates to WMS-IV subtests.

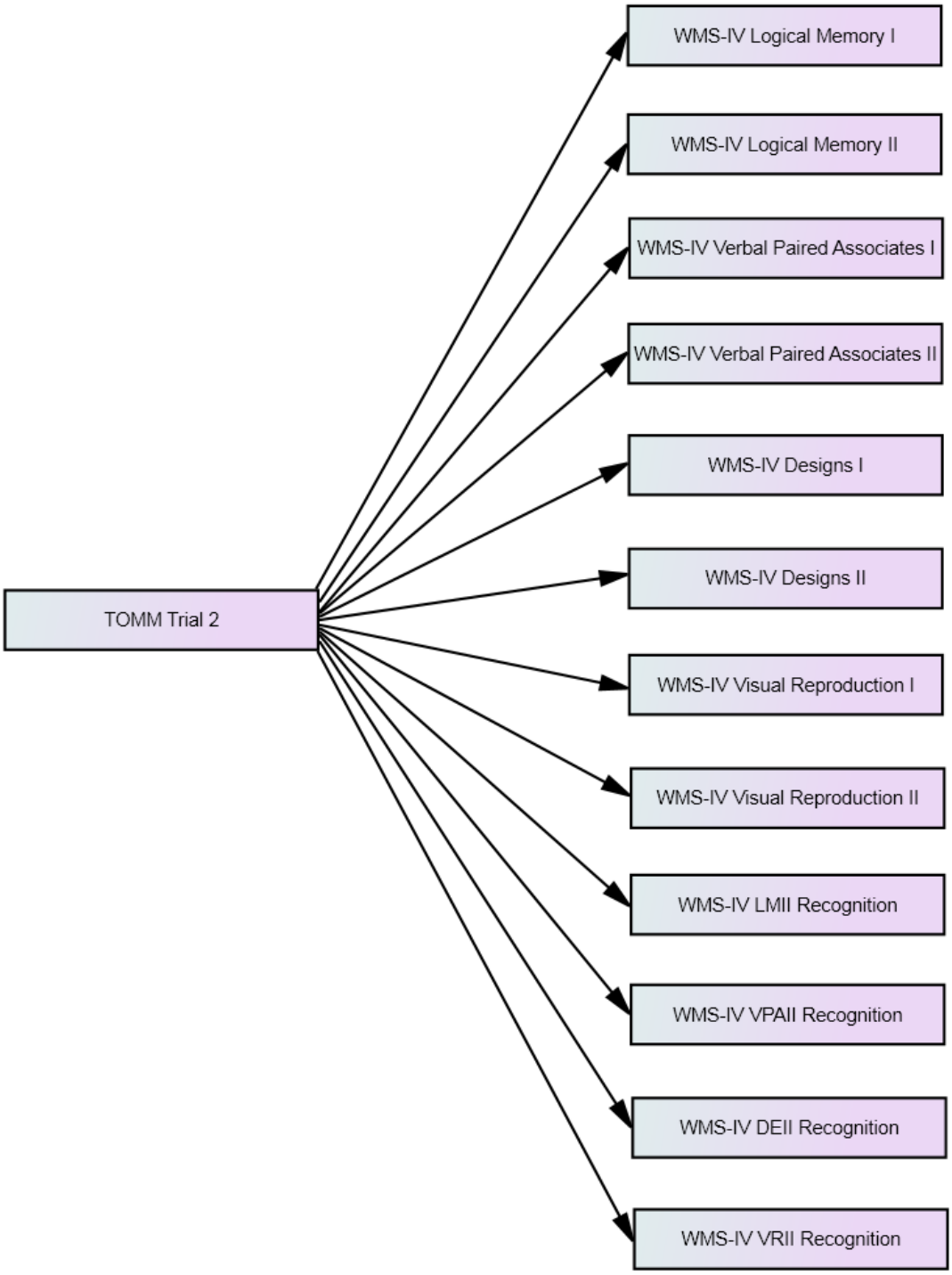


Figure 3: Regression model exploring how TOMM Trial 1 plus Trial 2 relate to WMS-IV subtests.

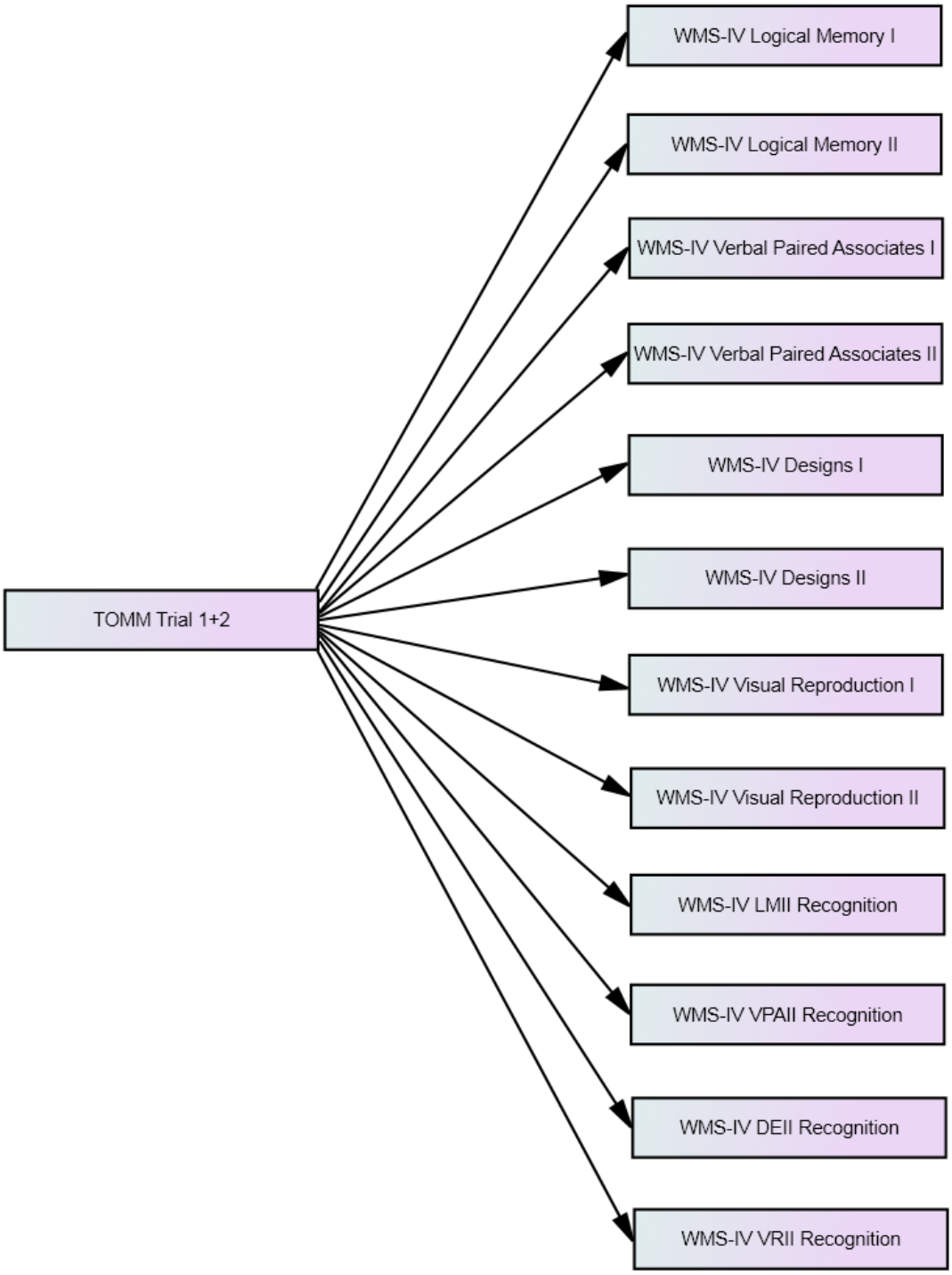


Figure 4: Regression model exploring how WMT Immediate Recall relates to WMS-IV subtests.

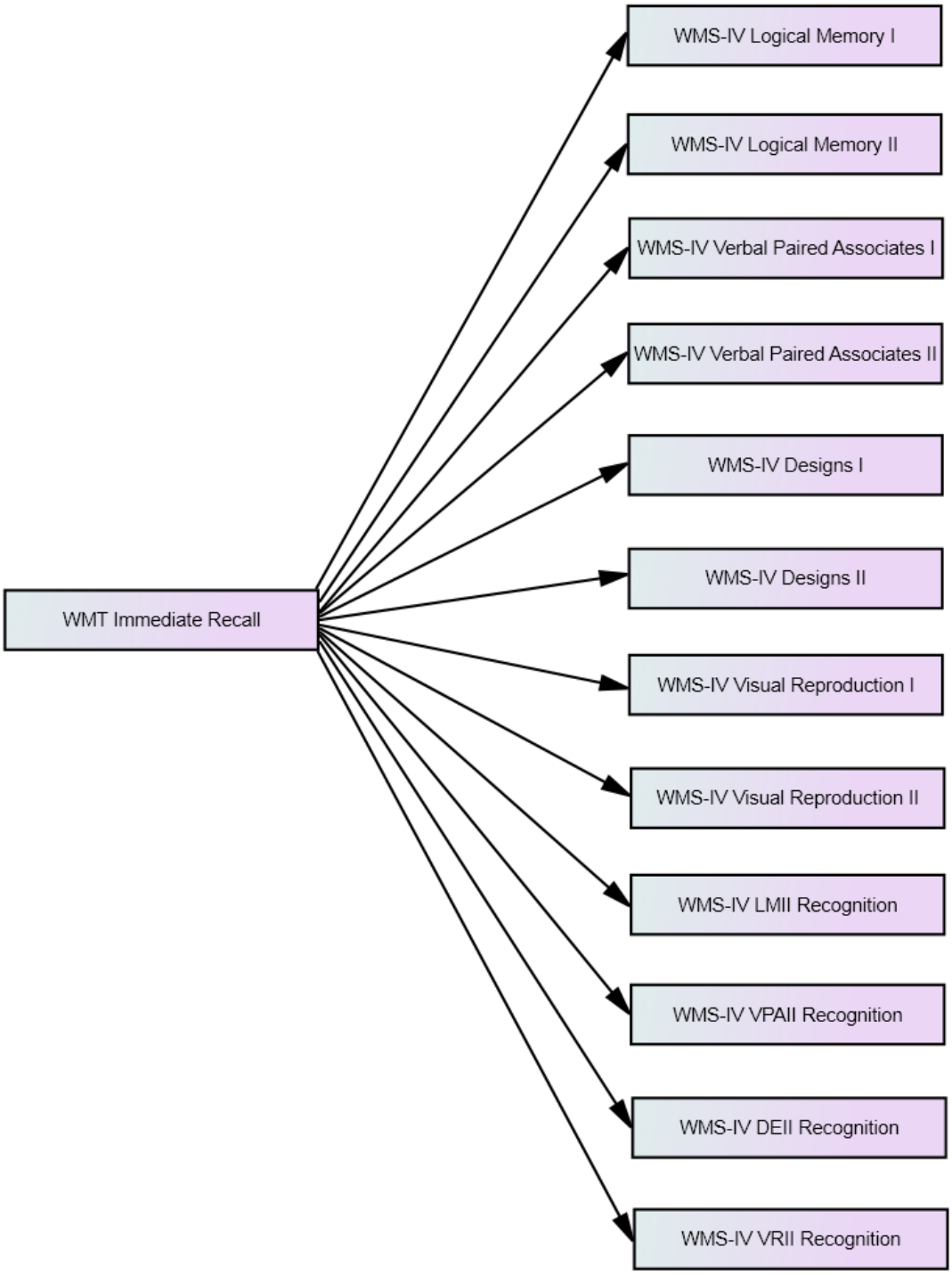


Figure 5: Regression model exploring how WMT Delayed Recall relates to WMS-IV subtests.

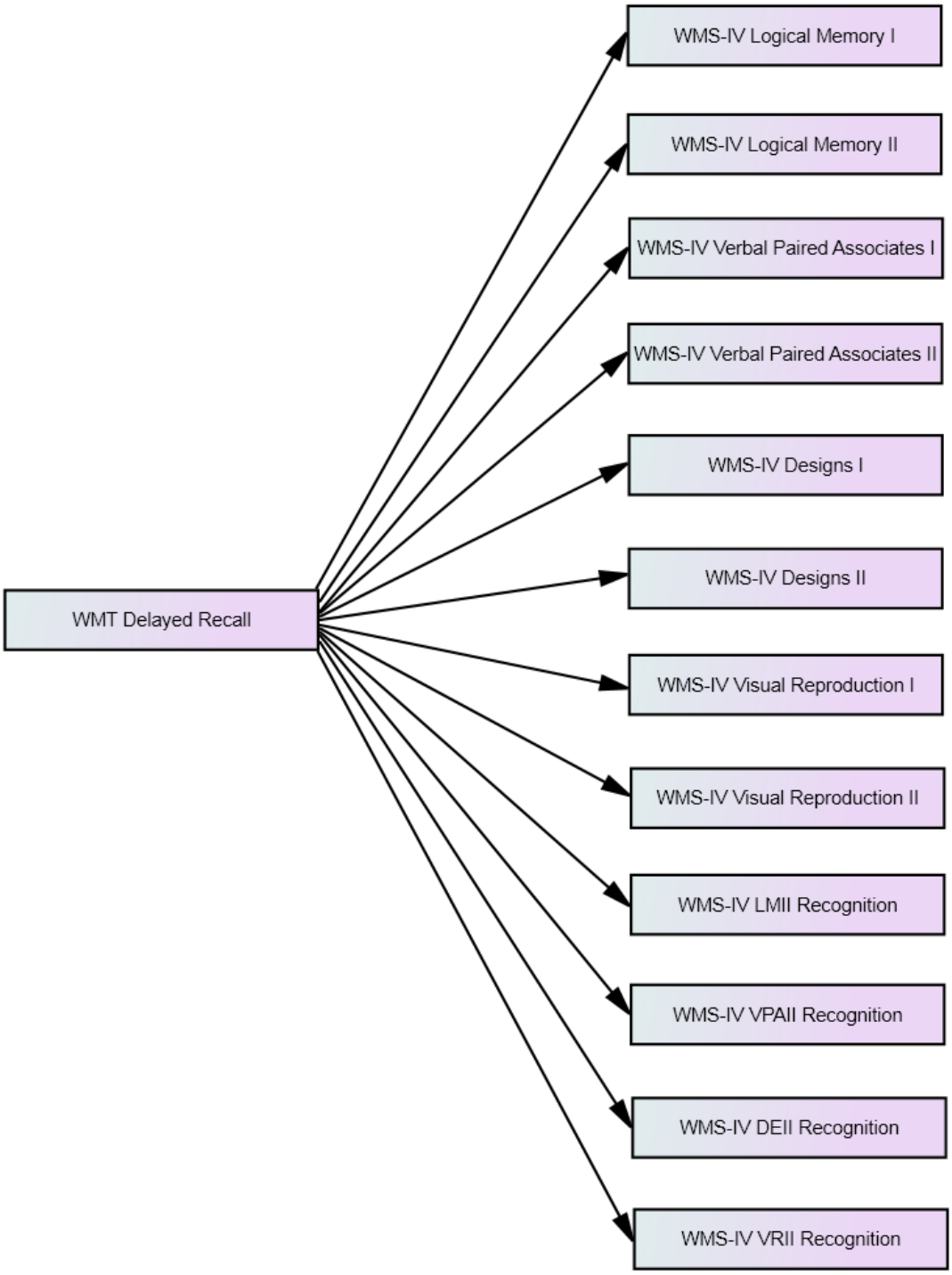


Figure 6: Regression model exploring how WMT Consistency relates to WMS-IV subtests.

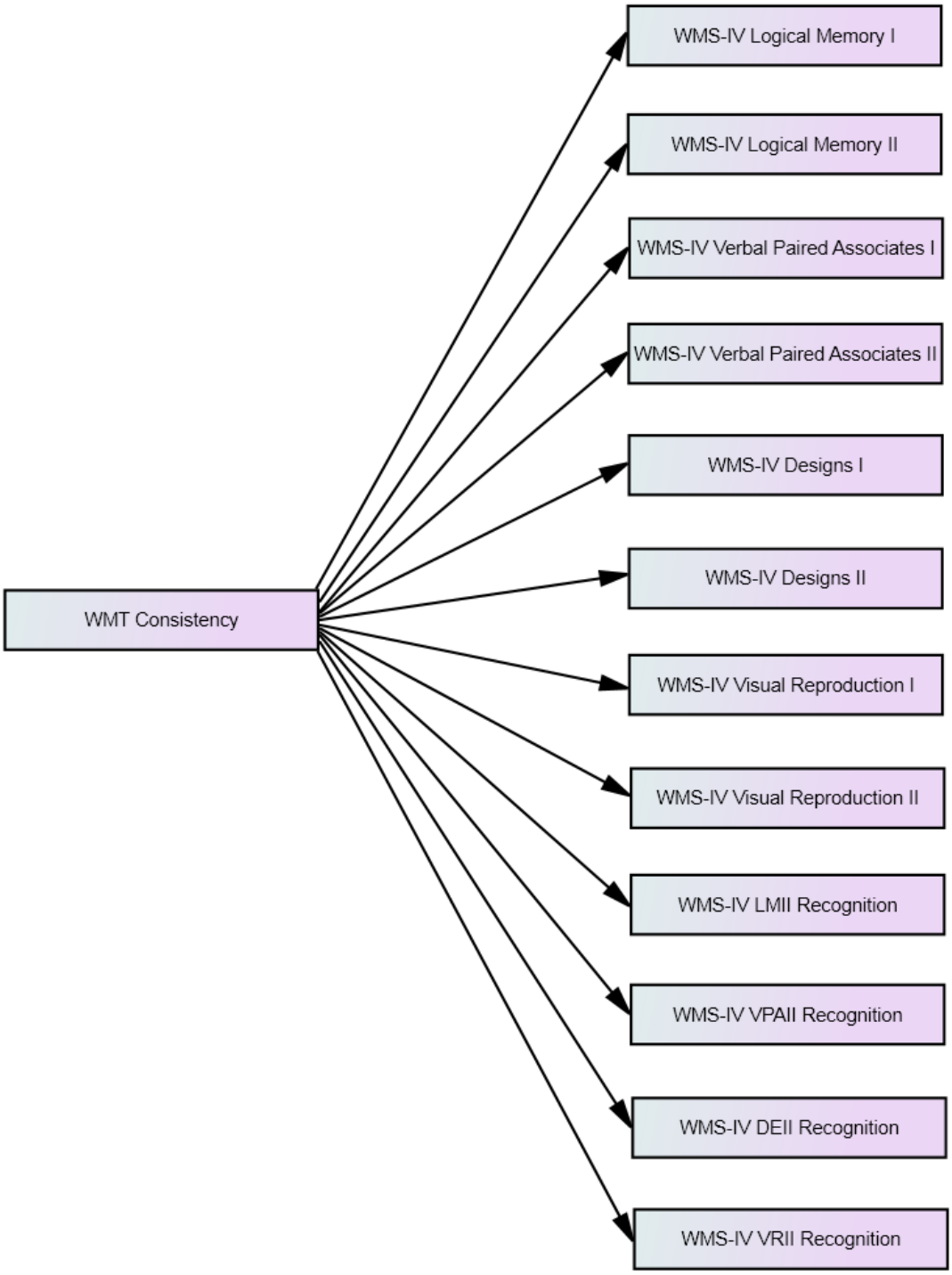


Figure 7: Regression model exploring how WMT Multiple Choice relates to WMS-IV subtests.

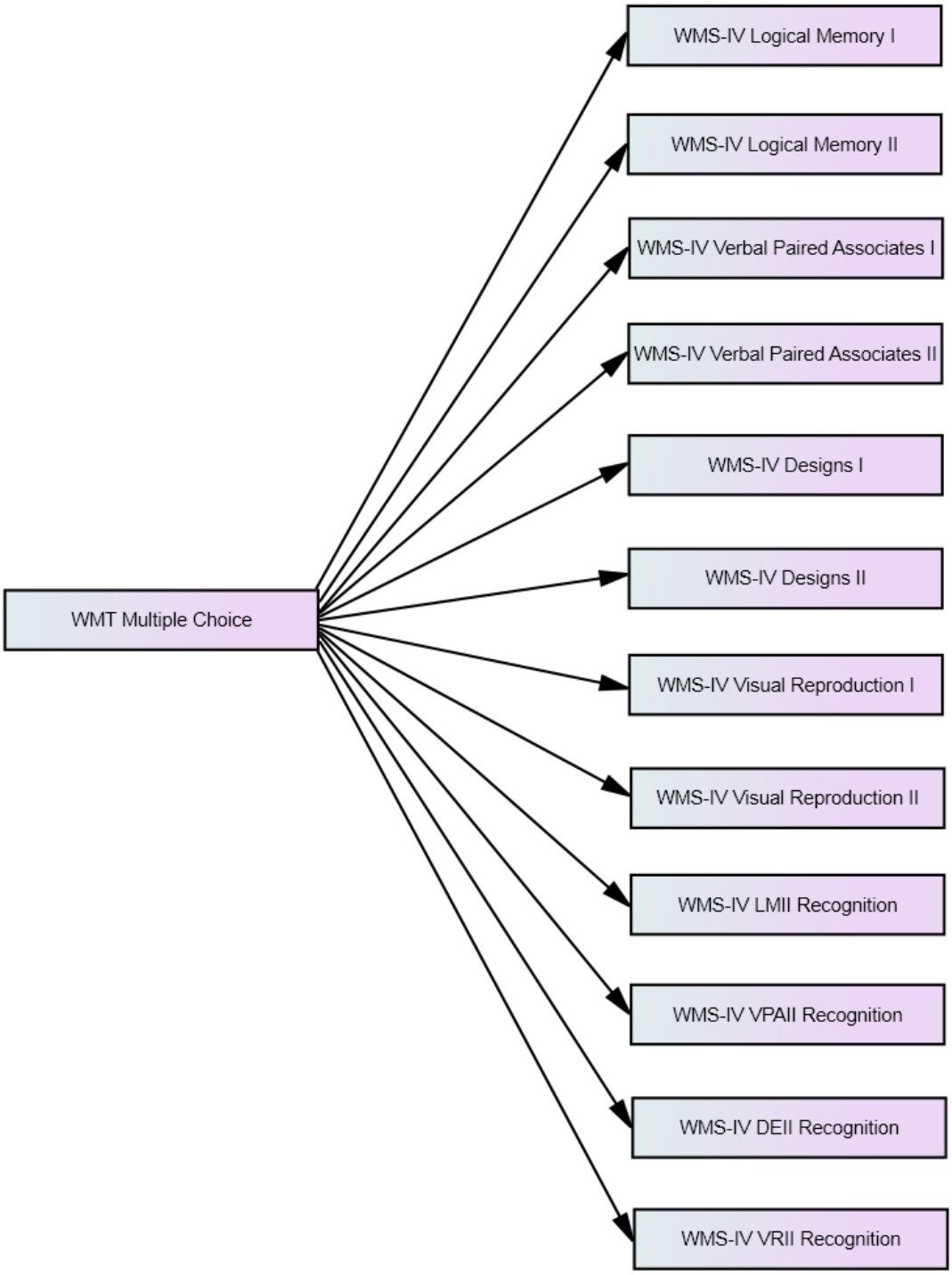


Figure 8: Regression model exploring how WMT Paired Associates relates to WMS-IV subtests.

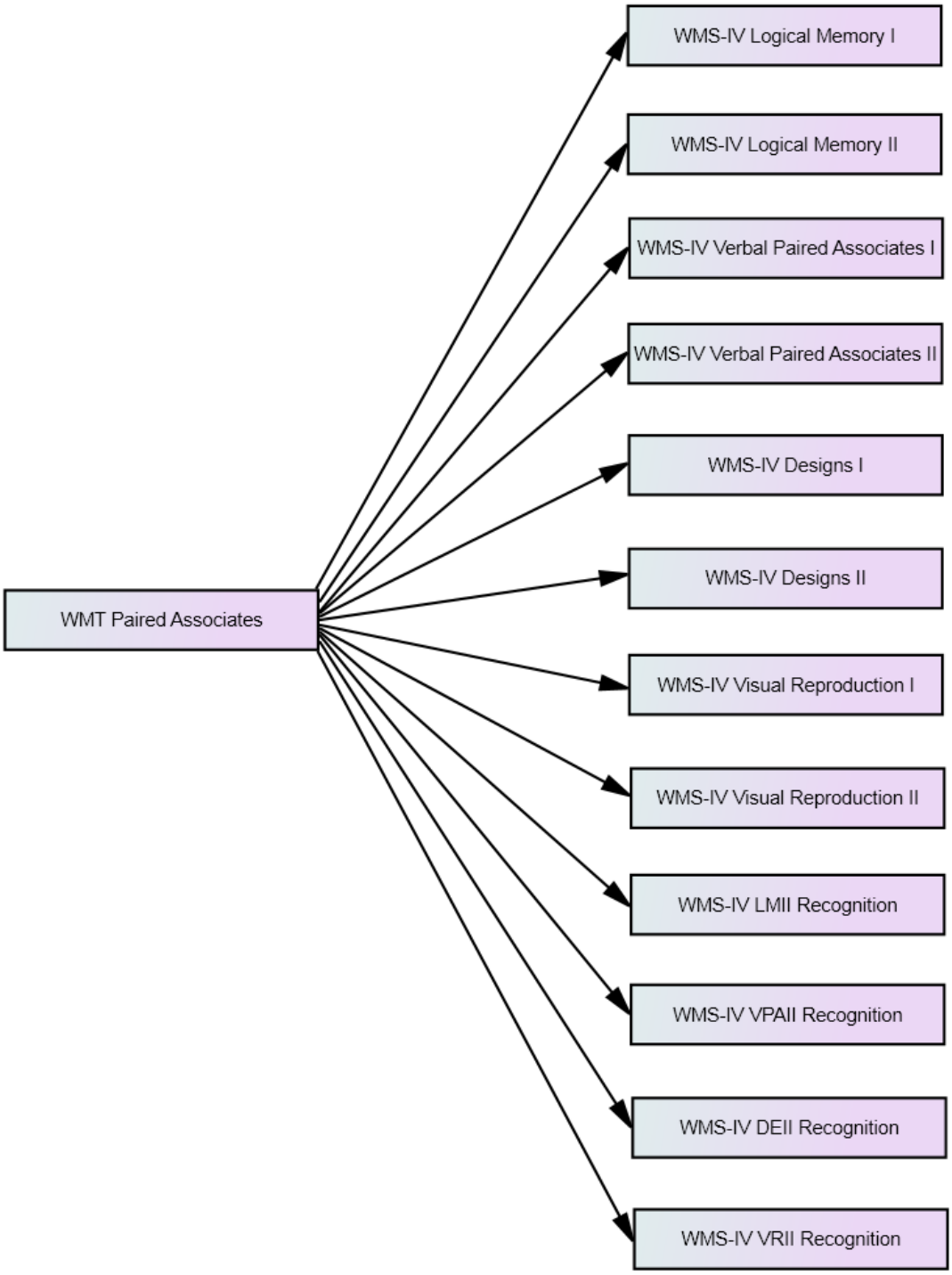


Figure 9: Regression model exploring how WMT Free Recall relates to WMS-IV subtests.

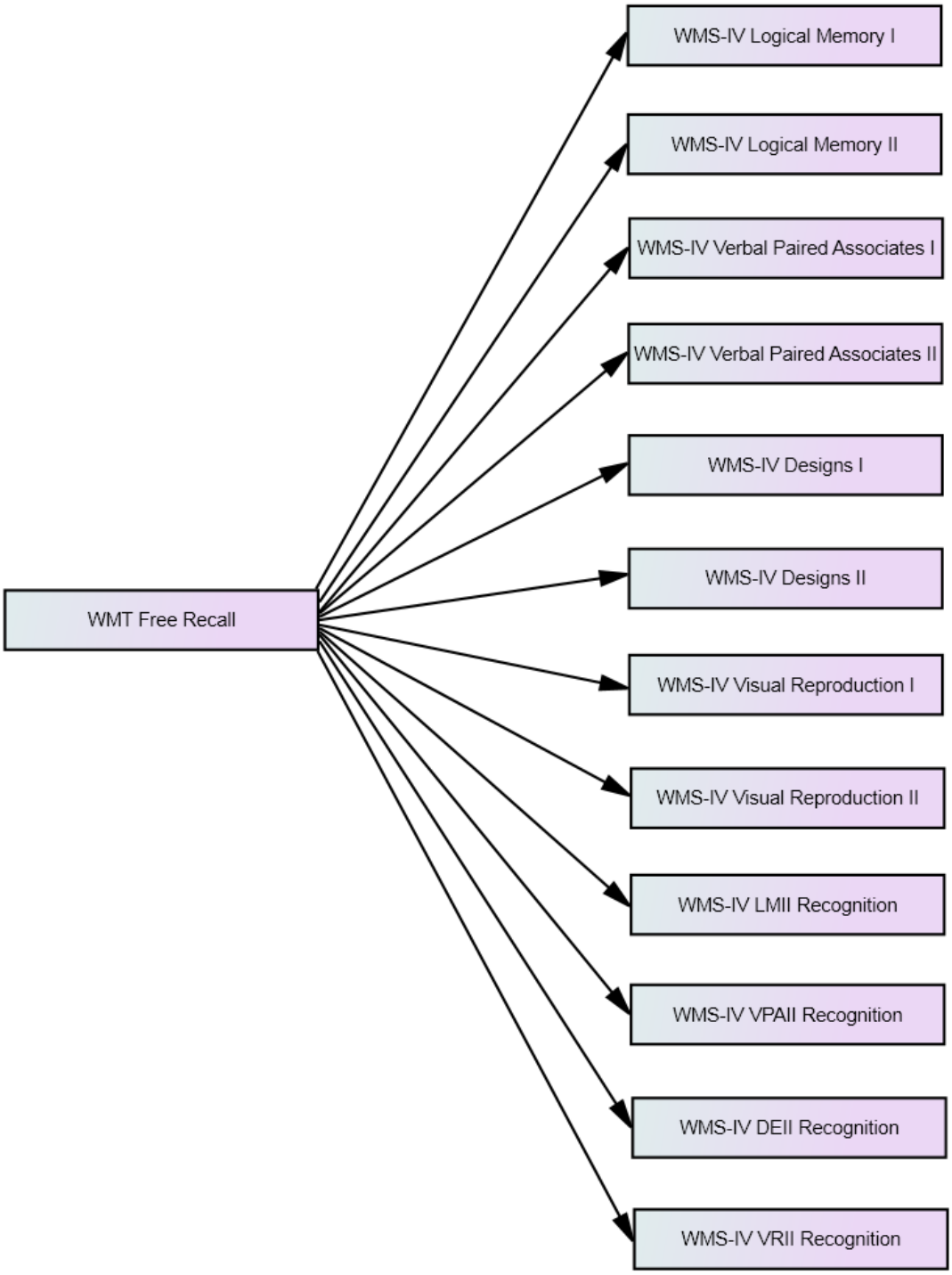


Figure 10: Regression model exploring how WMT LD Free Recall relates to WMS-IV subtests.

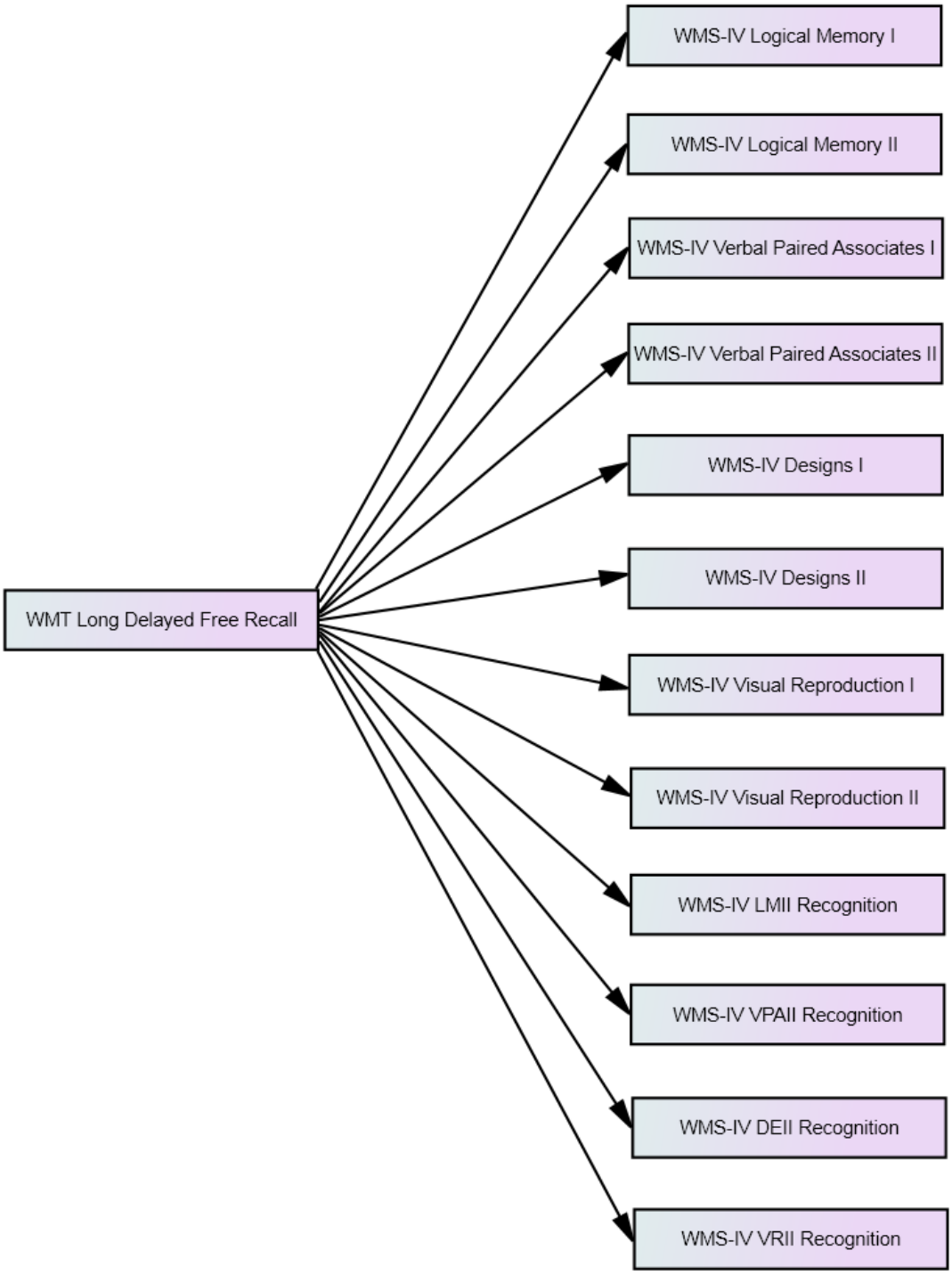


Figure 11: Regression model exploring how TOMM Trial 1 relates to WMS-IV composites.

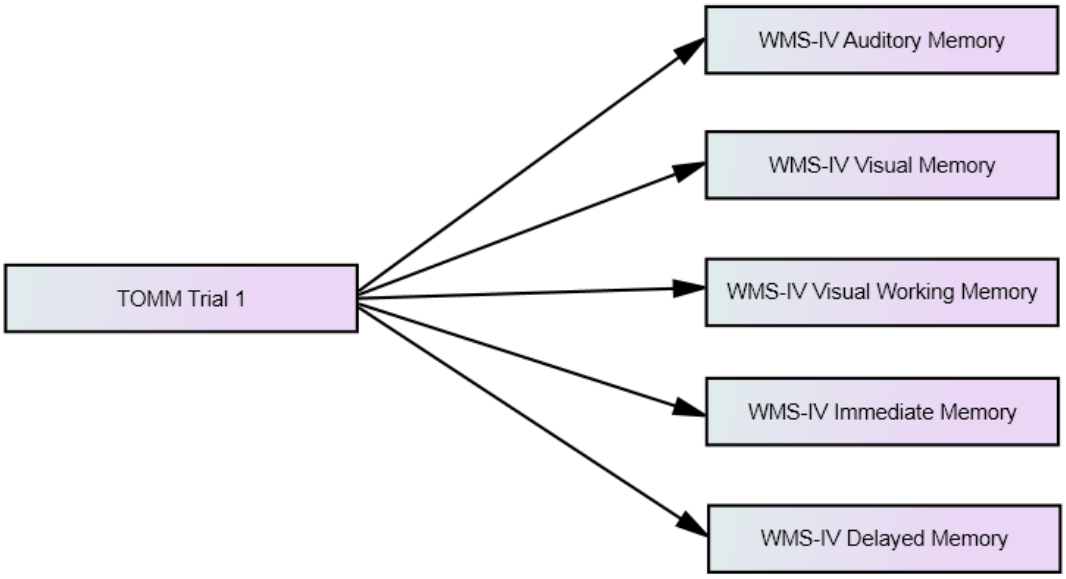


Figure 12: Regression model exploring how TOMM Trial 2 relates to WMS-IV composites.

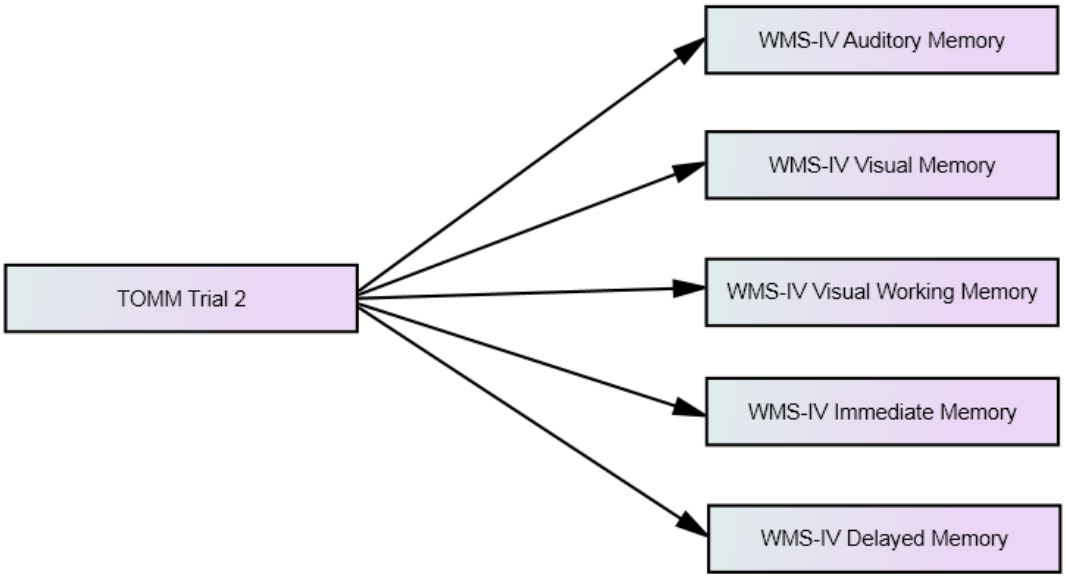


Figure 13: Regression model exploring how TOMM Trial 1 plus 2 relates to WMS-IV composites.

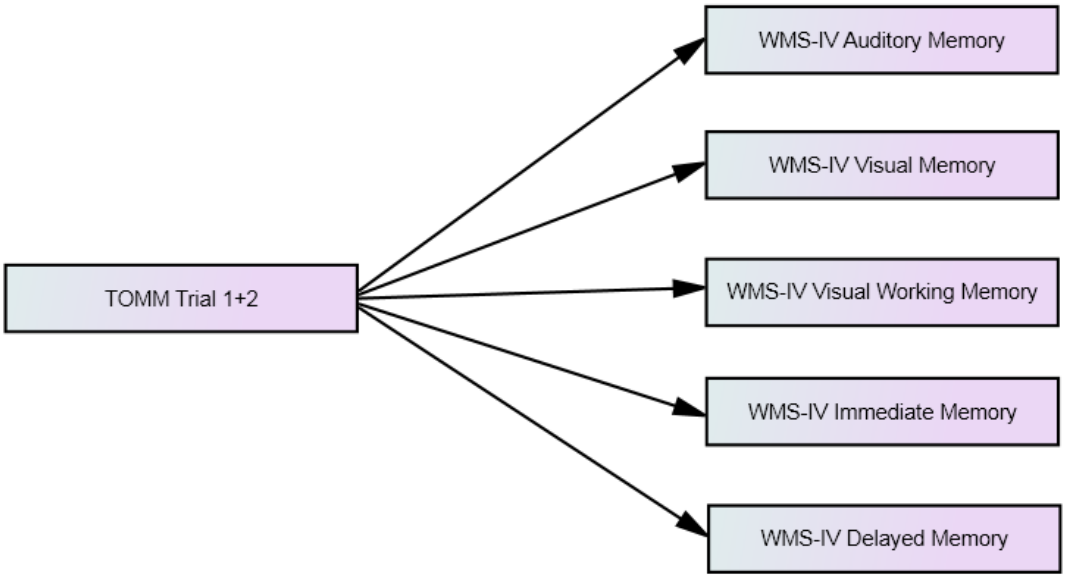


Figure 14: Regression model exploring how WMT Immediate Recall relates to WMS-IV composites.

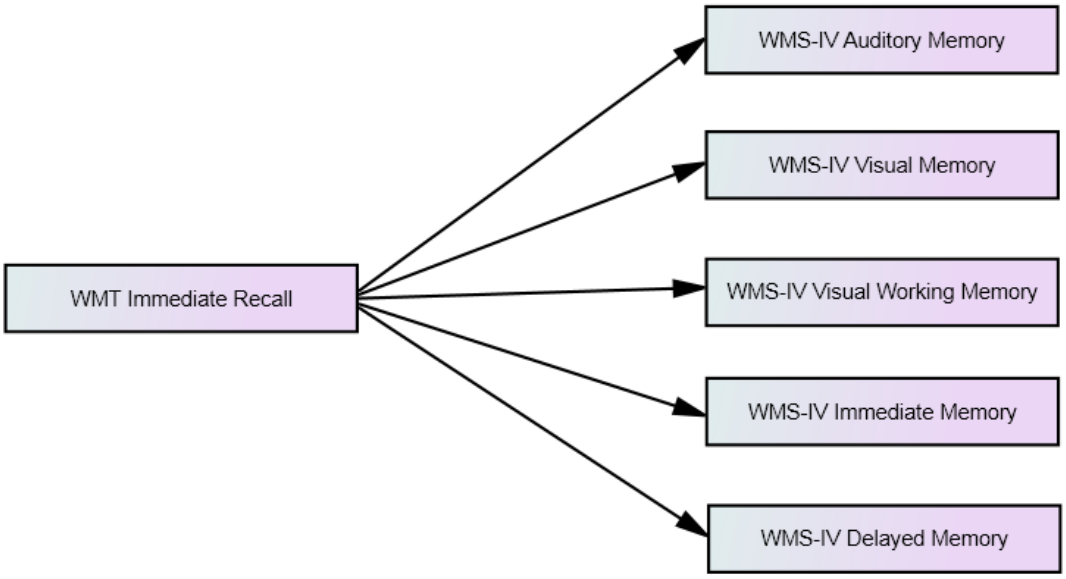


Figure 15: Regression model exploring how WMT Delayed Recall relates to WMS-IV composites.

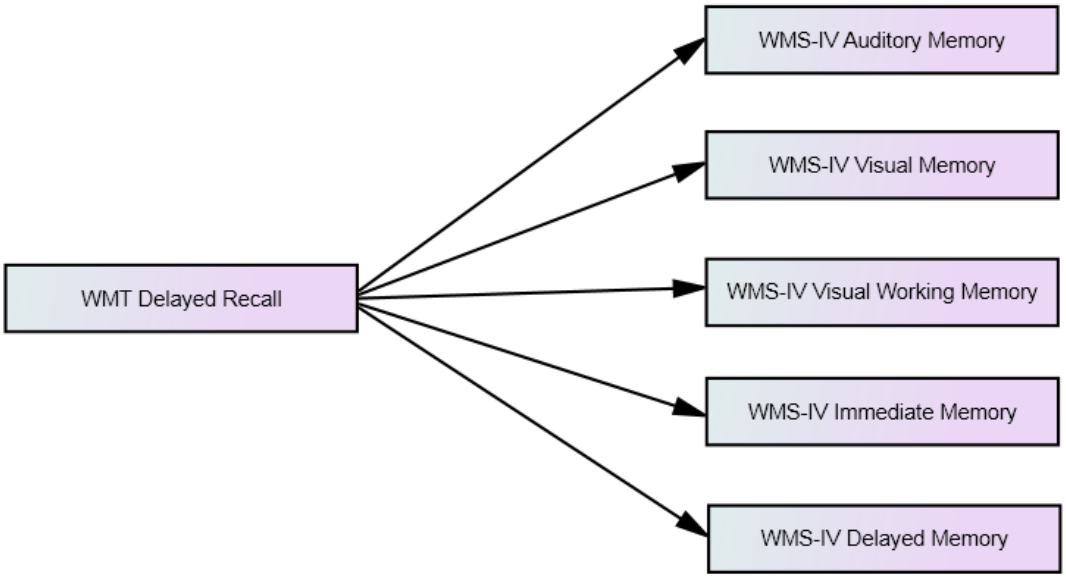


Figure 16: Regression model exploring how WMT Consistency relates to WMS-IV composites.

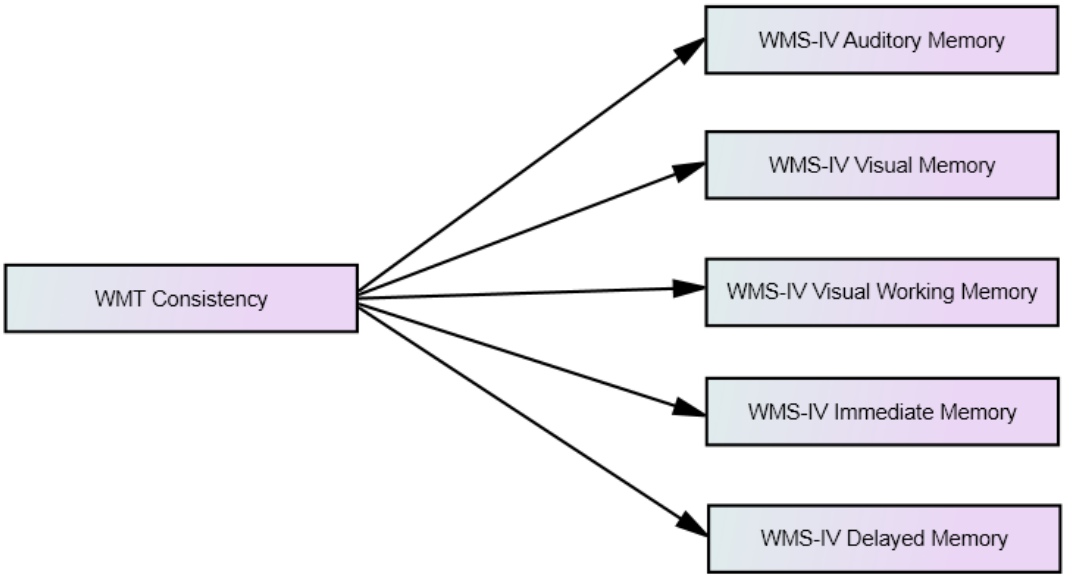


Figure 17: Regression model exploring how WMT Multiple Choice relates to WMS-IV composites.

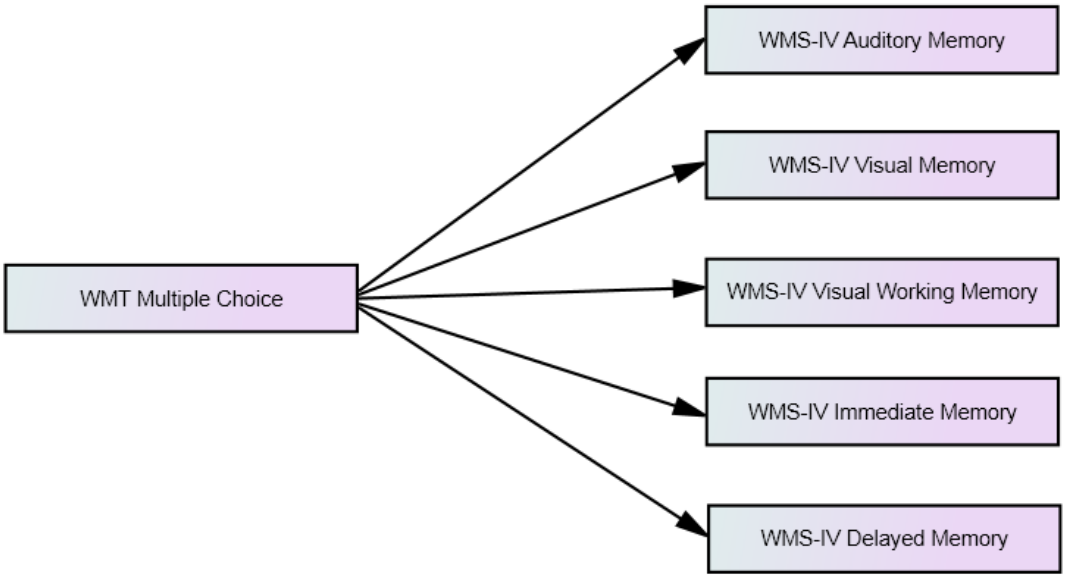


Figure 18: Regression model exploring how WMT Paired Associates relates to WMS-IV composites.

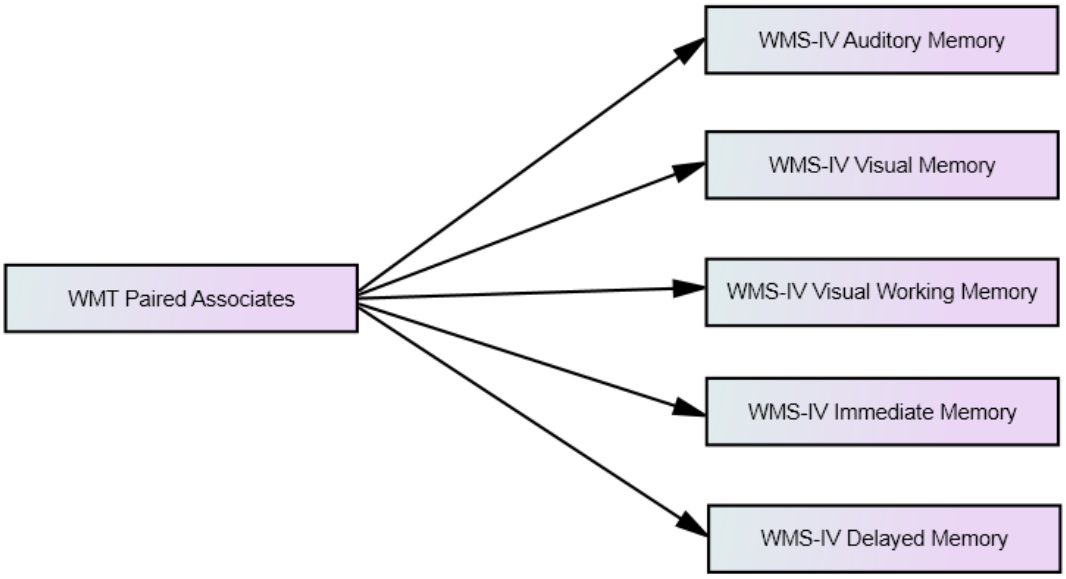


Figure 19: Regression model exploring how WMT Free Recall relates to WMS-IV composites.

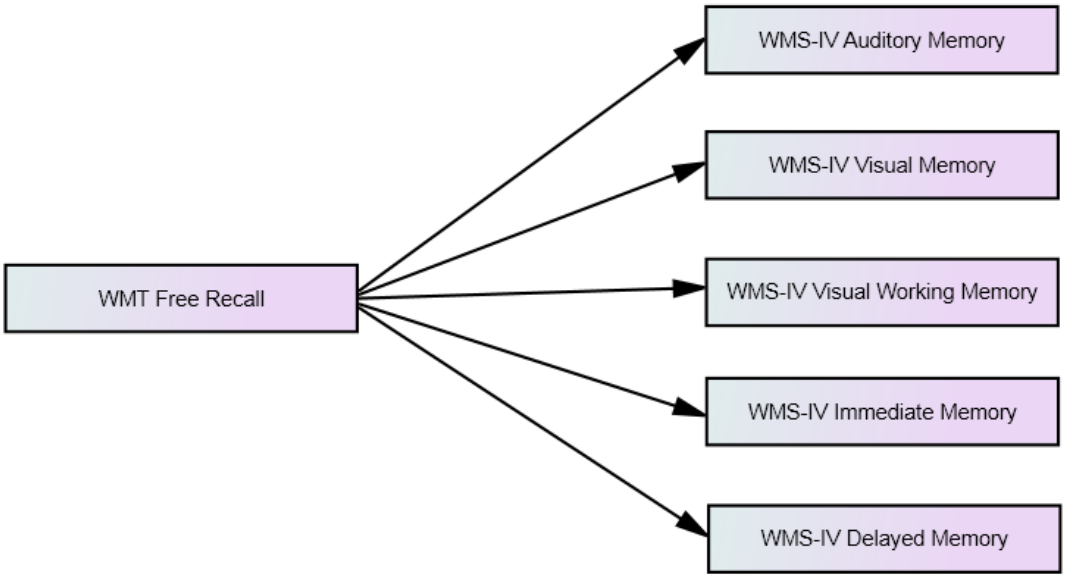
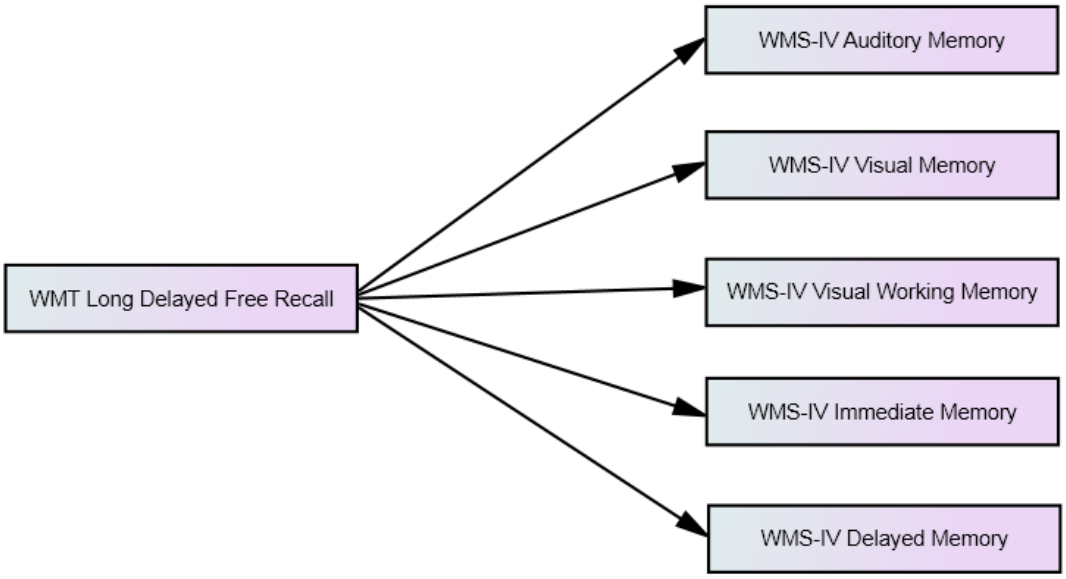


Figure 20: Regression model exploring how WMT LD Free Recall relates to WMS-IV composites.



Chapter IV

RESULTS

Results

The purpose of this study was to examine the relationship between measures of memory and measures of symptom validity testing. Specifically, this study was designed to explore the possibility that data from a widely used test of memory could be used to predict performance on common symptoms validity tests. Measures utilized in this study included data from the *Wechsler Memory Scale, Fourth Edition* (WMS-IV; Wechsler, 2009), the *Word Memory Test* (WMT; Green, 2003), and the *Test of Memory Malingering* (TOMM; Tombaugh, 1996). In total, seven research questions exploring relationships between memory measures and symptom validity measures were addressed.

This chapter consisted of a description of the sample and a summary of how the results of statistical tests addressed the research questions, and conclusions.

Description of the Sample. The sample consisted of 46 college students. Participants were required to be at least 18 years of age and enrolled in a psychology class at a medium-sized Midwestern university. Participants were awarded 4 hours of research participation credit for attending the study and were given the option to withdraw from the study at any time while still retaining full credit for participation. In total, 1 participant withdrew from the study which left 45 usable sets of data. Table 1 presents demographic characteristics of participants in the present study.

Table 1: *Demographic characteristics for the sample in the present study*

Variable	*n* = 46
Mean age (SD)	19.91 (1.62)
Gender	
Female	54%
Male	46%
Ethnicity	
Caucasian/White	72%
Multiracial/Mixed	11%
African American/Black	9%
Asian	4%
Hispanic/Latino	4%
Handedness	
Right	83%
Left	17%
Medical	
On one or more medications/supplements	54%
Not on medications or supplements	46%
One or more physical/mental diagnoses	35%
No physical or mental diagnoses	65%
ADHD diagnosis	9%
No ADHD diagnosis	81%
LD diagnosis	4%

No LD diagnosis		96%
TBI diagnosis		4%
No TBI diagnosis		96%
Educational Attribute Means*		
GPA (SD)	(*n* = 34)	2.80 (0.51)
SAT (SD)	(*n* = 26)	1621.81 (191.29)
ACT (SD)	(*n* = 20)	22.00 (3.51)
Years of education (SD)	(*n* = 46)	13.00 (1.30)
Mother's years of education (SD)	(*n* = 45)	14.59 (2.38)
Father's years of education (SD)	(*n* = 44)	14.49 (2.36)

* = Specific educational attributes were not reported by some participants. All other variables *n* = 46.

Summary of the Results

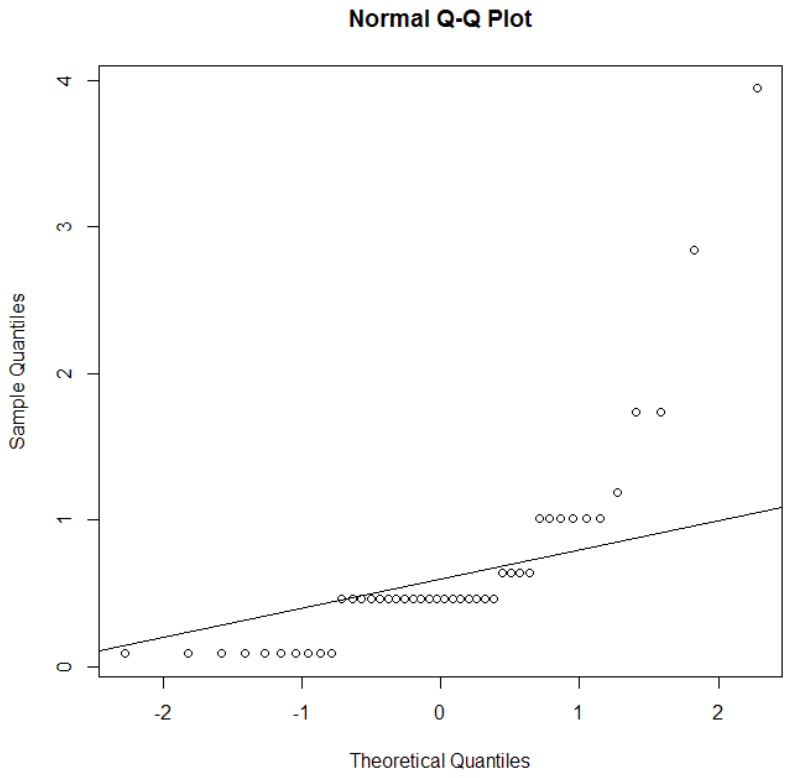
The application of the results to the research questions will now be discussed.

Regression Assumptions

Valid usage of parametric tests, such as linear regression, require data to adhere to statistical assumptions. Linear regression analyses require data to have multivariate normality, no significant multicollinearity, no auto-correlation, homoscedasticity, and a linear relationship.

Data adherence to assumptions was examined by using Q-Q Plots, quantitative information from the Mardia tests of multivariate skew and kurtosis in R (Mardia, 1970), and mean and standard deviation data. WMS-IV variables appeared free from restricted ranges; however, some degree of skew and exponential distribution was evidenced in WMS-IV variables. This was evidenced by Q-Q Plots (see figure 1) as well as mean and standard deviation data (see Table 23 on page 125). Figure 1 presents Verbal Paired Associates II, a WMS-IV subtest, which has data patterns representative of the other administered WMS-IV subtests.

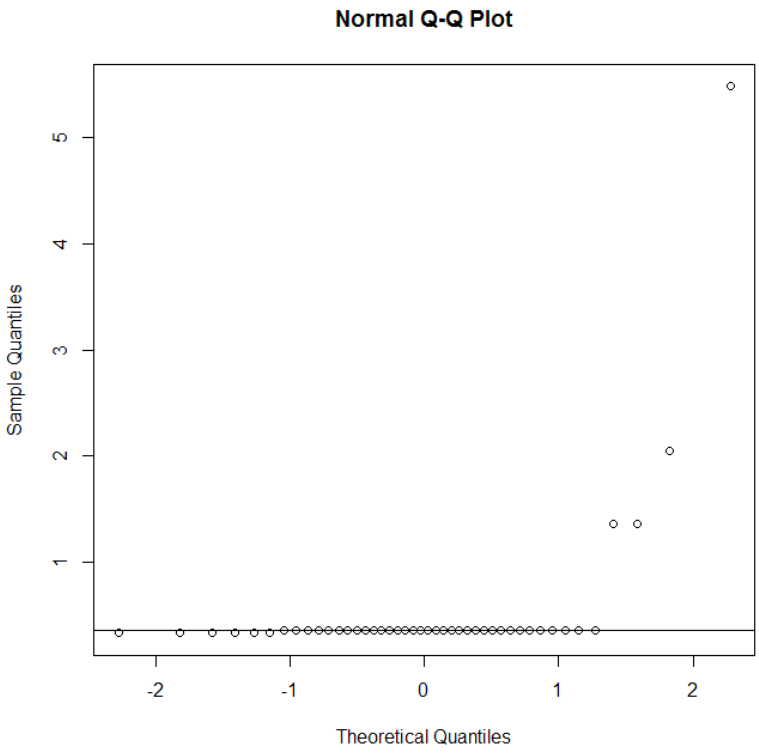
Figure 1: *Q-Q Plot for Verbal Paired Associates II (VPA2)*



Non-linear data transformations may correct for observed non-linear data patterns; however, some predictor variables had data distributions that may not be manipulated sufficiently for usage with parametric tests. Data for both TOMM trials had severely restricted ranges. TOMM Trial 1 had a mean score of 48.98 and a standard deviation of 1.31. For TOMM Trial 2, the sample had a mean score of 49.98 with a standard deviation of 0.16.

Some WMT predictor variables, specifically Multiple Choice (MC), Paired Associates (PR), Free Recall (FR), and Long Delayed Free Recall (LDFR) approached a normal and linear distribution; however, Immediate Recall (IR), Delayed Recall (DR), and Consistency (CNS) showed a curvilinear distribution and restricted ranges comparable to that observed with TOMM variables. Restricted ranges were evidenced by standard deviation and range data (see Table 23 on page 125). Data distribution abnormalities were also apparent on Q-Q Plots such as the plot presented for DR (see Figure 2).

Figure 2: *Q-Q Plot for WMT Delayed Recall (DR)*



Regression Results

R_1 How much variance in the WMS-IV subtests does the TOMM account for?

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on TOMM Trial 1. TOMM Trial 1 data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,21) = 1.2745$, $p < 0.2959$), with an R^2 of 0.041523. Statistical significance for the 20

analyses was set at the 0.0025 level rather than the 0.05 level due to using a Bonferroni correction (Dunn, 1961) to control for Type I error. Table 2 presents a regression model predicting WMS-IV subtest scores from TOMM Trial 1.

Table 2: *Regression model predicting WMS-IV subtest scores from TOMM Trial 1*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
TOMM1	0.50735	16	0.041523	0.2959	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	-13.0723		15.3759	-0.850	0.401
LM1 ~TOMM1	0.5007		0.3135	1.597	0.119
(Intercept)	4.9880		14.3085	0.349	0.729
LM2 ~ TOMM1	0.1205		0.2918	0.413	0.682
(Intercept)	23.0361		14.0065	1.645	0.109
VPA1 ~ TOMM1	-0.2503		0.2856	-0.877	0.387
(Intercept)	13.14458		11.72159	1.121	0.270
VPA2 ~ TOMM1	-0.03838		0.23901	-0.161	0.873
(Intercept)	-10.2771		17.6315	-0.583	0.564
DE1 ~ TOMM1	0.4378		0.3595	1.218	0.231
(Intercept)	-13.3133		17.6709	-0.753	0.456
DE2 ~ TOMM1	0.5029		0.3603	1.396	0.171

(Intercept)	24.4096	18.1380	1.346	0.187
VR1 ~ TOMM1	-0.2816	0.3698	-0.761	0.451
(Intercept)	15.74699	19.56104	0.805	0.426
VR2 ~ TOMM1	-0.09951	0.39886	-0.249	0.804
(Intercept)	12.1325	17.0220	0.713	0.481
SA ~ TOMM1	-0.0290	0.3471	-0.084	0.934
(Intercept)	1.7470	14.0719	0.124	0.902
SP ~ TOMM1	0.1968	0.2869	0.686	0.497
(Intercept)	-7.193	11.917	-0.604	0.550
LM2Rec ~ TOMM1	0.224	0.243	0.922	0.363
(Intercept)	2.19277	9.07457	0.242	0.81
VPA2Rec ~ TOMM1	0.03525	0.18504	0.191	0.85
(Intercept)	7.56627	15.30311	0.494	0.624
DE1Con ~ TOMM1	0.07809	0.31204	0.250	0.804
(Intercept)	17.8072	21.1693	0.841	0.406
DE1Spa ~ TOMM1	-0.1464	0.4317	-0.339	0.737

(Intercept)	-1.6265	17.0530	-0.095	0.925
DE2Con ~TOMM1	0.2651	0.3477	0.762	0.451
(Intercept)	6.95181	20.21590	0.344	0.733
DE2Spa ~ TOMM1	0.07452	0.41221	0.181	0.858

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on TOMM Trial 2. TOMM Trial 2 data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,21) = 0.89061$, $p < 0.5877$), with an R^2 of 0.031853. Table 3 presents a regression model predicting WMS-IV subtest scores from TOMM Trial 2.

Table 3: *Regression model predicting WMS-IV subtest scores from TOMM Trial 2*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
TOMM2	0.59575	16	0.031853	0.5877	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	-64.162		125.550	-0.511	0.612
LM1 ~ TOMM2	1.514		2.512	0.602	0.551

(Intercept)	-86.351	112.580	-0.767	0.448
LM2 ~ TOMM2	1.946	2.253	0.864	0.393
(Intercept)	74.243	111.756	0.664	0.511
VPA1 ~ TOMM2	-1.270	2.236	-0.568	0.574
(Intercept)	-2.2432	92.9627	-0.024	0.981
VPA2 ~ TOMM2	0.2703	1.8602	0.145	0.885
(Intercept)	207.027	138.891	1.491	0.145
DE1 ~ TOMM2	-3.919	2.779	-1.410	0.167
(Intercept)	96.432	143.178	0.674	0.505
DE2 ~ TOMM2	-1.703	2.865	-0.594	0.556
(Intercept)	184.838	142.057	1.301	0.201
VR1 ~ TOMM2	-3.486	2.843	-1.227	0.228
(Intercept)	325.568	146.133	2.228	0.0322
VR2 ~ TOMM2	-6.297	2.924	-2.154	0.0380
(Intercept)	128.216	133.576	0.96	0.344
SA ~ TOMM2	-2.351	2.673	-0.88	0.385

(Intercept)	42.4595	112.2031	0.378	0.707
SP ~ TOMM2	-0.6216	2.2452	-0.277	0.783
(Intercept)	168.568	91.585	1.841	0.0739
LM2Rec ~ TOMM2	-3.297	1.833	-1.799	0.0804
(Intercept)	-43.3514	71.5687	-0.606	0.548
VPA2Rec ~ TOMM2	0.9459	1.4321	0.661	0.513
(Intercept)	42.4595	121.3546	0.350	0.728
DE1Con ~ TOMM2	-0.6216	2.4284	-0.256	0.799
(Intercept)	337.486	159.079	2.122	0.0408
DE1Spa ~ TOMM2	-6.541	3.183	-2.055	0.0472
(Intercept)	43.7838	136.2168	0.321	0.750
DE2Con ~ TOMM2	-0.6486	2.7258	-0.238	0.813
(Intercept)	133.514	159.079	0.839	0.407
DE2Spa ~ TOMM2	-2.459	3.183	-0.773	0.445

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on a summation of performance on TOMM Trial 1 and TOMM Trial 2. A summation of TOMM Trial data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV

subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,21) = 1.3988$, $p < 0.2326$), with an R^2 of 0.04433. Table 4 presents a regression model predicting WMS-IV subtest scores from the sum of TOMM Trial 1 and 2.

Table 4: *Regression model predicting WMS-IV subtest scores from the sum of TOMM Trial 1 and 2*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
ALLT	0.48409	16	0.044330	0.2326	–
	Coefficient	Standard Error	Test Statistic	p-value	
(Intercept)	-41.4401	31.1666	-1.330	0.1920	
LM1 ~ ALLT	0.5345	0.3148	1.698	0.0982	
(Intercept)	-4.4673	29.0844	-0.154	0.879	
LM2 ~ ALLT	0.1552	0.2938	0.528	0.601	
(Intercept)	38.0735	28.4539	1.338	0.189	
VPA1 ~ ALLT	-0.2759	0.2874	-0.960	0.344	
(Intercept)	14.67695	23.86345	0.615	0.542	
VPA2 ~ ALLT	-0.03448	0.24103	-0.143	0.887	

(Intercept)	-26.3675	36.0853	-0.731	0.470
DE1 ~ ALLT	0.3793	0.3645	1.041	0.305
(Intercept)	-36.4510	36.0638	-1.011	0.319
DE2 ~ ALLT	0.4828	0.3643	1.325	0.193
(Intercept)	44.7432	36.7822	1.216	0.232
VR1 ~ ALLT	-0.3448	0.3715	-0.928	0.359
(Intercept)	31.3512	39.7084	0.790	0.435
VR2 ~ ALLT	-0.2069	0.4011	-0.516	0.609
(Intercept)	17.53811	34.63643	0.506	0.616
SA ~ ALLT	-0.06897	0.34984	-0.197	0.845
(Intercept)	-7.3811	28.6625	-0.258	0.798
SP ~ ALLT	0.1897	0.2895	0.655	0.517
(Intercept)	-13.2795	24.3785	-0.545	0.589
LM2Rec ~ ALLT	0.1724	0.2462	0.700	0.488
(Intercept)	-1.19964	18.46273	-0.065	0.949
VPA2Rec ~ ALLT	0.05172	0.18648	0.277	0.783

(Intercept)	4.56715	31.15891	0.147	0.884
DE1Con ~ ALLT	0.06897	0.31471	0.219	0.828
(Intercept)	36.2350	42.9516	0.844	0.404
DE1Spa ~ ALLT	-0.2586	0.4338	-0.596	0.555
(Intercept)	-14.2350	34.7327	-0.410	0.684
DE2Con ~ ALLT	0.2586	0.3508	0.737	0.466
(Intercept)	7.19147	41.16834	0.175	0.862
DE2Spa ~ ALLT	0.03448	0.41581	0.083	0.934

* = Significant at 0.0025 level

R_2 How much variance in the WMS-IV subtests does the WMT account for?

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Immediate Recall. WMT Immediate Recall data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 1.4183$, $p < 0.2137$), with an R^2 of 0.040743. Table 5 presents a regression model predicting WMS-IV subtest scores from WMT Immediate Recall.

Table 5: *Regression model predicting WMS-IV subtest scores from WMT Immediate Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
IR	0.514	16	0.040743	0.2137	–
		Coefficient	Standard Error	Test Statistic	p-value
(Intercept)		-4.0846	8.8501	-0.462	0.6470
LM1 ~ IR		0.1591	0.0902	1.764	0.0856
(Intercept)		-7.24667	7.55003	-0.960	0.343
LM2 ~ IR		0.18664	0.07695	2.425	0.020
(Intercept)		-0.92458	8.17709	-0.113	0.911
VPA1 ~ IR		0.11916	0.08334	1.430	0.161
(Intercept)		-5.30532	6.35921	-0.834	0.4092
VPA2 ~ IR		0.16808	0.06482	2.593	0.0133
(Intercept)		-14.76151	9.70418	-1.521	0.1363
DE1 ~ IR		0.26430	0.09891	2.672	0.0109
(Intercept)		-11.48917	9.54724	-1.203	0.2361
DE2 ~ IR		0.23166	0.09731	2.381	0.0223
(Intercept)		-10.1872	10.2123	-0.998	0.3247
VR1 ~ IR		0.2102	0.1041	2.019	0.0504

(Intercept)	-6.0548	10.6809	-0.567	0.574
VR2 ~ IR	0.1715	0.1089	1.575	0.123
(Intercept)	-16.98099	9.08632	-1.869	0.06917
SA ~ IR	0.28072	0.09261	3.031	0.00431
(Intercept)	-8.24419	7.46285	-1.105	0.2761
SP ~ IR	0.20155	0.07606	2.650	0.0116
(Intercept)	4.249327	7.068457	0.601	0.551
LM2Rec ~ IR	-0.004036	0.072045	-0.056	0.956
(Intercept)	7.66032	5.21018	1.470	0.150
VPA2Rec ~ IR	-0.03809	0.05310	-0.717	0.478
(Intercept)	-3.30842	8.34416	-0.396	0.6939
DE1Con ~ IR	0.14945	0.08505	1.757	0.0867
(Intercept)	-11.0742	11.9226	-0.929	0.3587
DE1Spa ~ IR	0.2212	0.1215	1.820	0.0764
(Intercept)	-2.89247	9.38552	-0.308	0.760
DE2Con ~ IR	0.14521	0.09566	1.518	0.137

(Intercept)	-5.8334	11.1587	-0.523	0.604
DE2Spa ~ IR	0.1663	0.1137	1.462	0.152

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Delayed Recall. WMT Delayed Recall data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 1.9006$, $p < 0.07522$), with an R^2 of 0.049869. Table 6 presents a regression model predicting WMS-IV subtest scores from WMT Delayed Recall.

Table 6: *Regression model predicting WMS-IV subtest scores from WMT Delayed Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
DR	0.4411	16	0.049869	0.07522	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	-7.9510		13.3328	-0.596	0.554
LM1~ DR	0.1964		0.1345	1.460	0.152
(Intercept)	-9.8916		11.5834	-0.854	0.3983
LM2 ~ DR	0.2113		0.1168	1.809	0.0782

(Intercept)	-8.1058	12.1190	-0.669	0.508
VPA1 ~ DR	0.1903	0.1222	1.557	0.128
(Intercept)	-16.92547	9.21137	-1.837	0.07377
VPA2 ~ DR	0.28352	0.09292	3.051	0.00409
(Intercept)	-23.3708	14.7109	-1.589	0.1202
DE1 ~ DR	0.3483	0.1484	2.347	0.0241
(Intercept)	-10.7304	14.8009	-0.725	0.473
DE2 ~ DR	0.2215	0.1493	1.484	0.146
(Intercept)	-18.9064	15.2743	-1.238	0.2232
VR1 ~ DR	0.2959	0.1541	1.920	0.0621
(Intercept)	-11.1493	16.0213	-0.696	0.491
VR2 ~ DR	0.2210	0.1616	1.368	0.179
(Intercept)	-38.7227	12.8015	-3.025	0.004386
SA ~ DR	0.4971	0.1291	3.849	0.000428
(Intercept)	-8.2554	11.6475	-0.709	0.4827
SP ~ DR	0.1995	0.1175	1.698	0.0975

(Intercept)	9.45805	10.48636	0.902	0.373
LM2Rec ~ DR	-0.05655	0.10578	-0.535	0.596
(Intercept)	9.31596	7.76062	1.200	0.237
VPA2Rec ~ DR	-0.05438	0.07828	-0.695	0.491
(Intercept)	1.71378	12.81372	0.134	0.894
DE1Con ~ DR	0.09715	0.12925	0.752	0.457
(Intercept)	-22.4302	17.7168	-1.266	0.2130
DE1Spa ~ DR	0.3334	0.1787	1.866	0.0696
(Intercept)	6.88764	14.36337	0.48	0.634
DE2Con ~ DR	0.04494	0.14488	0.31	0.758
(Intercept)	-9.0592	16.7743	-0.540	0.592
DE2Spa ~ DR	0.1970	0.1692	1.164	0.251

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Consistency. WMT Consistency data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired

Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 1.2062$, $p < 0.3305$), with an R^2 of 0.03621. Table 7 presents a regression model predicting WMS-IV subtest scores from WMT Consistency.

Table 7: *Regression model predicting WMS-IV subtest scores from WMT Consistency*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
CNS	0.55427	16	0.03621	0.3305	–
	Coefficient	Standard Error	Test Statistic	p-value	
(Intercept)	-2.33733	9.12155	-0.256	0.799	
LM1 ~ CNS	0.14164	0.09321	1.520	0.137	
(Intercept)	-6.39738	7.78087	-0.822	0.4160	
LM2 ~ CNS	0.17842	0.07951	2.244	0.0306	
(Intercept)	-2.97684	8.27554	-0.360	0.721	
VPA1 ~ CNS	0.14045	0.08456	1.661	0.105	
(Intercept)	-5.71015	6.48745	-0.880	0.384	
VPA2 ~ CNS	0.17264	0.06629	2.604	0.013	
(Intercept)	-14.5389	9.9575	-1.460	0.1523	
DE1 ~CNS	0.2627	0.1017	2.582	0.0137	
(Intercept)	-7.7529	9.9770	-0.777	0.4418	

DE2 ~ CNS	0.1940	0.1019	1.903	0.0644
(Intercept)	-11.2747	10.3901	-1.085	0.2845
VR1 ~ CNS	0.2218	0.1062	2.089	0.0433
(Intercept)	-7.0328	10.8773	-0.647	0.522
VR2 ~CNS	0.1819	0.1111	1.637	0.110
(Intercept)	-15.22283	9.44767	-1.611	0.11518
SA ~ CNS	0.26344	0.09654	2.729	0.00948
(Intercept)	-4.51015	7.86703	-0.573	0.5697
SP ~ CNS	0.16386	0.08039	2.038	0.0483
(Intercept)	3.030460	7.214634	0.420	0.677
LM2Rec ~ CNS	0.008419	0.073721	0.114	0.910
(Intercept)	9.38185	5.28170	1.776	0.0835
VPA2Rec ~ CNS	-0.05579	0.05397	-1.034	0.3076
(Intercept)	-4.31345	8.48545	-0.508	0.6141
DE1Con ~ CNS	0.16010	0.08671	1.846	0.0724
(Intercept)	-9.5957	12.2567	-0.783	0.438

DE1Spa ~ CNS	0.2066	0.1252	1.650	0.107
(Intercept)	-1.56121	9.64051	-0.162	0.872
DE2Con ~ CNS	0.13196	0.09851	1.340	0.188
(Intercept)	-0.5703	11.5644	-0.049	0.961
DE2Spa ~ CNS	0.1128	0.1182	0.955	0.345

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Multiple Choice. WMT Multiple Choice data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). A statistically significant regression equation was found ($F(16,24) = 2.2879$, $p < 0.03247$), with an R^2 of 0.056252. This suggests that WMT Multiple Choice scores account for over 5% of the variance in WMS-IV subtest scores. Table 8 presents a regression model predicting WMS-IV subtest scores from WMT Multiple Choice.

Table 8: *Regression model predicting WMS-IV subtest scores from WMT Multiple Choice*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
MC	0.396	16	0.056252	0.03247	–
		Coefficient	Standard Error	Test Statistic	p-value
(Intercept)		7.13225	4.13726	1.724	0.0926

LM1 ~ MC	0.04683	0.04404	1.063	0.2942
(Intercept)	2.24491	3.41430	0.658	0.5147
LM2 ~ MC	0.09412	0.03634	2.590	0.0134
(Intercept)	-0.22492	3.39653	-0.066	0.9475
VPA1 ~ MC	0.11740	0.03616	3.247	0.0024
(Intercept)	1.84639	2.76143	0.669	0.50766
VPA2 ~ MC	0.09969	0.02940	3.391	0.00161
(Intercept)	-0.83895	4.41447	-0.190	0.85026
DE1 ~ MC	0.12813	0.04699	2.727	0.00953
(Intercept)	1.65694	4.40187	0.376	0.7086
DE2 ~ MC	0.10223	0.04686	2.182	0.0352
(Intercept) -	2.65556	4.42401	-0.600	0.55180
VR1 ~ MC	0.13973	0.04709	2.967	0.00511
(Intercept)	2.78056	4.86059	0.572	0.571
VR2 ~ MC	0.08527	0.05174	1.648	0.107
(Intercept)	-2.30270	4.12086	-0.559	0.57950

SA ~ MC	0.13726	0.04387	3.129	0.00331
(Intercept)	5.93005	3.58901	1.652	0.107
SP ~ MC	0.05968	0.03820	1.562	0.126
(Intercept)	5.62030	3.21332	1.749	0.0881
LM2Rec ~ MC	-0.01889	0.03421	-0.552	0.5840
(Intercept)	7.51509	2.32271	3.235	0.00248
VPA2Rec ~ MC	-0.03836	0.02473	-1.552	0.12885
(Intercept)	6.06720	3.86377	1.570	0.124
DE1Con ~ MC	0.05639	0.04113	1.371	0.178
(Intercept)	-0.25764	5.39118	-0.048	0.9621
DE1Spa ~ MC	0.11618	0.05739	2.024	0.0498
(Intercept)	6.96885	4.35138	1.602	0.117
DE2Con ~ MC	0.04675	0.04632	1.009	0.319
(Intercept)	0.91183	4.99929	0.182	0.8562
DE2Spa ~ MC	0.10212	0.05322	1.919	0.0623

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Paired Associates. WMT Paired Associates data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 2.0906$, $p < 0.04972$), with an R^2 of 0.053092. Table 9 presents a regression model predicting WMS-IV subtest scores from WMT Paired Associates.

Table 9: *Regression model predicting WMS-IV subtest scores from WMT Paired Associates*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
PR	0.41776	16	0.053092	0.04972	–

	Coefficient	Standard Error	Test Statistic	p-value
(Intercept)	6.78485	3.81282	1.779	0.083
LM1 ~ PR	0.05061	0.04061	1.246	0.220
(Intercept)	3.35859	3.19276	1.052	0.2993
LM2 ~ PR	0.08232	0.03401	2.421	0.0202
(Intercept)	0.4616	3.1357	0.147	0.88372
VPA1 ~ PR	0.1102	0.0334	3.300	0.00208
(Intercept)	1.08384	2.41644	0.449	0.656256

VPA2 ~ PR	0.10798	0.02574	4.195	0.000152
(Intercept)	-0.57609	4.04409	-0.142	0.88746
DE1 ~ PR	0.12549	0.04307	2.913	0.00589
(Intercept)	1.30875	4.01443	0.326	0.7462
DE2 ~ PR	0.10609	0.04276	2.481	0.0175
(Intercept)	-3.4340	3.9525	-0.869	0.39026
VR1 ~ PR	0.1482	0.0421	3.521	0.00111
(Intercept)	0.20370	4.33639	0.047	0.9628
VR2 ~ PR	0.11296	0.04619	2.446	0.0191
(Intercept)	-1.84007	3.77704	-0.487	0.62886
SA ~ PR	0.13249	0.04023	3.293	0.00211
(Intercept)	4.33468	3.22683	1.343	0.1869
SP ~ PR	0.07684	0.03437	2.236	0.0312
(Intercept)	7.45185	2.93185	2.542	0.0151
LM2Rec ~ PR	-0.03852	0.03123	-1.233	0.2248
(Intercept)	5.72593	2.19823	2.605	0.0129

VPA2Rec ~ PR	-0.01926	0.02341	-0.823	0.4158
(Intercept)	5.48182	3.54130	1.548	0.130
DE1Con ~ PR	0.06273	0.03772	1.663	0.104
(Intercept)	-0.98687	4.90747	-0.201	0.8417
DE1Spa ~ PR	0.12414	0.05227	2.375	0.0226
(Intercept)	5.48182	3.97312	1.380	0.176
DE2Con ~ PR	0.06273	0.04232	1.482	0.146
(Intercept)	0.78855	4.58812	0.172	0.8644
DE2Spa ~ PR	0.10357	0.04887	2.119	0.0405

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Free Recall. WMT Free Recall data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 1.4305$, $p < 0.2083$), with an R^2 of 0.040994. Table 10 presents a regression model predicting WMS-IV subtest scores from WMT Free Recall.

Table 10: *Regression model predicting WMS-IV subtest scores from WMT Free Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
FR	0.51185	16	0.040994	0.2083	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	8.17510		1.77628	4.602	4.35e-05
LM1 ~ FR	0.05146		0.02678	1.921	0.062
(Intercept)	7.01307		1.49742	4.683	3.38e-05
LM2 ~ FR	0.06223		0.02258	2.756	0.00885
(Intercept)	7.79230		1.62426	4.797	2.37e-05
VPA1 ~ FR	0.04570		0.02449	1.866	0.0696
(Intercept)	8.10456		1.29774	6.245	2.36e-07
VPA2 ~ FR	0.04728		0.01957	2.416	0.0205
(Intercept)	8.64702		2.09342	4.131	0.000185
DE1 ~ FR	0.03854		0.03156	1.221	0.229426
(Intercept)	9.94626		2.05424	4.842	2.06e-05
DE2 ~ FR	0.01963		0.03097	0.634	0.53
(Intercept)	6.35273		2.06443	3.077	0.00381
VR1 ~ FR	0.06263		0.03113	2.012	0.05114

(Intercept)	5.31657	2.04007	2.606	0.01290
VR2 ~ FR	0.08387	0.03076	2.727	0.00953
(Intercept)	6.96875	1.95567	3.563	0.000985
SA ~ FR	0.05501	0.02949	1.866	0.069613
(Intercept)	10.04419	1.62053	6.198	2.75e-07
SP ~ FR	0.02264	0.02443	0.926	0.36
(Intercept)	5.45942	1.40408	3.888	0.000382
LM2Rec ~ FR	-0.02476	0.02117	-1.170	0.249266
(Intercept)	4.70102	1.05221	4.468	6.6e-05
VPA2Rec ~ FR	-0.01194	0.01586	-0.752	0.456
(Intercept)	10.04581	1.73882	5.777	1.06e-06
DE1Con ~ FR	0.01998	0.02622	0.762	0.451
(Intercept)	8.10403	2.47588	3.273	0.00223
DE1Spa ~ FR	0.03864	0.03733	1.035	0.30704
(Intercept)	10.44751	1.94640	5.368	3.9e-06
DE2Con ~ FR	0.01378	0.02935	0.470	0.641

(Intercept)	9.42410	2.30969	4.08	0.000215
DE2Spa ~ FR	0.01603	0.03482	0.46	0.647937

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on subtest scores from the WMS-IV based on performance on WMT Long Delayed Free Recall. WMT Long Delayed Free Recall data was used as a predictor for WMS-IV subtest score dependent variables. WMS-IV subtest score data consisted of Logical Memory II (LM2), Verbal Paired Associates I (VPA1), Verbal Paired Associates II (VPA2), Designs I (DE1), Designs II (DE2), Visual Reproduction I (VR1), Visual Reproduction II (VR2), Spatial Addition (SA), Symbol Span (SP), Logical Memory II Recognition (LM2Recog), Verbal Paired Associates II Recognition (VPA2Recog), Designs I Content (DE1Con), Designs I Spatial (DE1Spa), Designs II Content (DE2Con), and Designs II Spatial (DE2 Spa). No statistically significant regression equation was found ($F(16,24) = 0.86505$, $p < 0.6108$), with an R^2 of 0.028057. Table 11 presents a regression model predicting WMS-IV subtest scores from WMT Long Delayed Free Recall.

Table 11: *Regression model predicting WMS-IV subtest scores from WMT Long Delayed Free Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
LDFR	0.63424	16	0.028057	0.6108	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	8.31561		1.83903	4.522	5.59e-05
LM1 ~ LDFR	0.04843		0.02728	1.775	0.0836
(Intercept)	6.83511		1.53637	4.449	7e-05
LM2 ~ LDFR	0.06384		0.02279	2.801	0.00788

(Intercept)	7.16635	1.64218	4.364	9.09e-05
VPA1 ~ LDFR	0.05439	0.02436	2.233	0.0314
(Intercept)	8.53825	1.36523	6.254	2.3e-07
VPA2 ~ LDFR	0.03989	0.02025	1.969	0.056
(Intercept)	8.62497	2.15531	4.002	0.000272
DE1 ~ LDFR	0.03820	0.03197	1.195	0.239356
(Intercept)	10.01438	2.11502	4.735	2.88e-05
DE2 ~ LDFR	0.01826	0.03137	0.582	0.564
(Intercept)	7.25224	2.17055	3.341	0.00185
VR1 ~ LDFR	0.04792	0.03220	1.488	0.14476
(Intercept)	7.08088	2.20980	3.204	0.0027
VR2 ~ LDFR	0.05569	0.03278	1.699	0.0973
(Intercept)	7.15197	2.02550	3.531	0.00108
SA ~ LDFR	0.05128	0.03005	1.707	0.09582
(Intercept)	11.050657	1.683692	6.563	8.57e-08
SP ~ LDFR	0.006993	0.024976	0.280	0.781

(Intercept)	4.69981	1.46304	3.212	0.00264
LM2Rec ~ LDFR	-0.01282	0.02170	-0.591	0.55811
(Intercept)	4.302898	1.088568	3.953	0.000315
VPA2Rec ~ LDFR	-0.005698	0.016148	-0.353	0.726093
(Intercept)	9.71753	1.78242	5.452	2.98e-06
DE1Con ~ LDFR	0.02461	0.02644	0.931	0.358
(Intercept)	7.77215	2.53979	3.060	0.00399
DE1Spa ~ LDFR	0.04299	0.03768	1.141	0.26076
(Intercept)	10.58933	2.00427	5.283	5.1e-06
DE2Con ~ LDFR	0.01140	0.02973	0.383	0.704
(Intercept)	9.24974	2.37428	3.896	0.000373
DE2Spa ~ LDFR	0.01839	0.03522	0.522	0.604547

* = Significant at 0.0025 level

R_3 How much variance in the WMS-IV composites does the TOMM account for?

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on TOMM Trial 1. TOMM Trial 1 data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index

(VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,32) = 0.058231$, $p < 0.9976$), with an R^2 of 0.001811. Table 12 presents a regression model predicting WMS-IV index scores from TOMM Trial 1.

Table 12: *Regression model predicting WMS-IV index scores from TOMM Trial 1*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
TOMM1	0.99098	5	0.001811	0.9975	–

	Coefficient	Standard Error	Test Statistic	p-value
(Intercept)	84.181	62.775	1.341	0.188
AMI ~ TOMM1	0.452	1.280	0.353	0.726
(Intercept)	61.398	92.849	0.661	0.513
VMI ~ TOMM1	0.913	1.893	0.482	0.633
(Intercept)	81.349	76.929	1.057	0.297
VWMI ~ TOMM1	0.506	1.569	0.323	0.749
(Intercept)	74.4337	77.0973	0.965	0.341
IMI ~ TOMM1	0.6627	1.5721	0.422	0.676
(Intercept)	66.7349	79.1975	0.843	0.405
DMI ~ TOMM1	0.8358	1.6149	0.518	0.608

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on TOMM Trial 2. TOMM Trial 2 data was used as a predictor for

WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,32) = 1.338$, $p < 0.2736$), with an R^2 of 0.037257. Table 13 presents a regression model predicting WMS-IV index scores from TOMM Trial 2.

Table 13: *Regression model predicting WMS-IV index scores from TOMM Trial 2*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
TOMM2	0.82709	5	0.037257	0.2736	–

	Coefficient	Standard Error	Test Statistic	p-value
(Intercept)	-116.514	497.305	-0.234	0.816
AMI ~ TOMM2	4.459	9.951	0.448	0.657
(Intercept)	1432.49	704.85	2.032	0.0495
VMI ~ TOMM2	-26.54	14.10	-1.882	0.0680
(Intercept)	559.973	606.256	0.924	0.362
VWMI ~ TOMM2	-9.081	12.131	-0.749	0.459
(Intercept)	726.86	604.15	1.203	0.237
IMI ~ TOMM2	-12.41	12.09	-1.026	0.312
(Intercept)	635.81	624.22	1.019	0.315
DMI ~ TOMM2	-10.57	12.49	-0.846	0.403

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the sum of TOMM Trial 1 and Trial 2. A summation of TOMM Trial data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,32) = 0.047738$, $p < 0.9985$), with an R^2 of 0.001484. Table 14 presents a regression model predicting WMS-IV index scores from the summation of TOMM Trial 1 and 2.

Table 14: *Regression model predicting WMS-IV index scores from a summation of TOMM Trial 1and 2*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
ALLT	0.9926	5	0.001484	0.9985	–
	Coefficient	Standard Error	Test Statistic	p-value	
(Intercept)	53.4283	127.7079	0.418	0.678	
AMI ~ ALLT	0.5345	1.2899	0.414	0.681	
(Intercept)	58.3648	189.4557	0.308	0.760	
VMI ~ ALLT	0.4828	1.9135	0.252	0.802	
(Intercept)	70.3131	156.7165	0.449	0.656	
VWMI ~ ALLT	0.3621	1.5829	0.229	0.820	
(Intercept)	60.8348	157.1465	0.387	0.701	
IMI ~ ALLT	0.4655	1.5872	0.293	0.771	

(Intercept)	41.1416	161.4405	0.255	0.800
DMI ~ ALLT	0.6724	1.6306	0.412	0.683

* = Significant at 0.0025 level

R_4 How much variance in the WMS-IV composites does the WMT account for?

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Immediate Recall score of the Word Memory Test (WMT). Immediate Recall data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 4.4651$, $p < 0.002988$), with an R^2 of 0.093967. Table 15 presents a regression model predicting WMS-IV index scores from WMT Immediate Recall.

Table 15: *Regression model predicting WMS-IV index scores from WMT Immediate Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
IR	0.61055	5	0.0939967	0.002988	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	13.5470		34.4651	0.393	0.6964
AMI ~ IR	0.9484		0.3513	2.700	.0102
(Intercept)	-28.4390		48.5135	-0.586	0.56111
VMI ~ IR	1.3658		0.4945	2.762	0.00871
(Intercept)	-37.5060		36.8370	-1.018	0.31488

VWMI ~ IR	1.4640	0.3755	3.899	0.00037
(Intercept)	-18.8146	40.5893	-0.464	0.64556
IMI ~ IR	1.2793	0.4137	3.092	0.00366
(Intercept)	-22.8618	39.4299	-0.580	0.56538
DMI ~ IR	1.3290	0.4019	3.307	0.00203

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Delayed Recall score of the WMT. Delayed Recall data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 4.0836$, $p < 0.005028$), with an R^2 of 0.087812. Table 16 presents a regression model predicting WMS-IV index scores from WMT Delayed Recall.

Table 16: *Regression model predicting WMS-IV index scores from WMT Delayed Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
DR	0.63157	5	0.087812	0.005028	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	-22.8194		51.9247	-0.439	0.6627
AMI ~ DR	1.3051		0.5238	2.492	0.0171
(Intercept)	-61.1117		74.3391	-0.822	0.4160

VMI ~ DR	1.6807	0.7499	2.241	0.0308
(Intercept)	-105.2151	55.1006	-1.910	0.063573
VWMI ~ DR	2.1314	0.5558	3.835	0.000447
(Intercept)	-65.265	61.562	-1.060	0.29559
IMI ~DR	1.734	0.621	2.793	0.00806
(Intercept)	-51.8615	61.3354	-0.846	0.4030
DMI ~ DR	1.6073	0.6187	2.598	0.0132

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Consistency score of the WMT. Consistency score data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 3.2457$, $p < 0.01641$), with an R^2 of 0.07336. Table 17 presents a regression model predicting WMS-IV index scores from WMT Consistency.

Table 17: *Regression model predicting WMS-IV index scores from WMT Consistency*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
CNS	0.68321	5	0.07336	0.01641	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	13.2823		35.2970	0.376	0.7087

AMI ~ CNS	0.9535	0.3607	2.644	0.0118
(Intercept)	-26.5166	49.8504	-0.532	0.5978
VMI ~ CNS	1.3495	0.5094	2.649	0.0116
(Intercept)	-21.0376	39.3795	-0.534	0.59622
VWMI ~ CNS	1.2992	0.4024	3.229	0.00252
(Intercept)	-21.3602	41.4394	-0.515	0.60914
IMI ~ CNS	1.3085	0.4234	3.090	0.00368
(Intercept)	-17.1216	40.9394	-0.418	0.67808
DMI ~ CNS	1.2736	0.4183	3.045	0.00416

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Multiple Choice score of the WMT. Multiple Choice score data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 3.9367$, $p < 0.006162$), with an R^2 of 0.085376. Table 18 presents a regression model predicting WMS-IV index scores from WMT Multiple Choice.

Table 18: *Regression model predicting WMS-IV index scores from WMT Multiple Choice*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025

MC	0.64005	5	0.085376	0.006162	–

	Coefficient	Standard Error	Test Statistic	p-value
(Intercept)	57.3505	15.2021	3.773	0.000537
AMI ~ MC	0.5256	0.1618	3.248 0	.002394
(Intercept)	40.0294	21.8032	1.836	0.07400
VMI ~ MC	0.6993	0.2321	3.013	0.00453
(Intercept)	50.2482	17.6743	2.843	0.00708
VWMI ~ MC	0.5960	0.1881	3.168	0.00298
(Intercept)	38.8080	17.5619	2.210	0.033058
IMI ~ MC	0.7246	0.1869	3.876	0.000396
(Intercept)	45.6556	17.7744	2.569	0.01415
DMI ~ MC	0.6603	0.1892	3.490	0.00122

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Paired Associates score of the WMT. Paired Associates data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). A statistically significant regression equation was found ($F(5,35) = 5.7812$, $p < 0.000543$), with an R^2 of 0.113446. The p value reached statistical significance at the 0.0025 (0.05/20) level. This suggests that

WMT Paired Associates scores account for over 11% of the variance in WMS-IV composite scores. Table 19 presents a regression model predicting WMS-IV index scores from WMT Paired Associates.

Table 19: *Regression model predicting WMS-IV index scores from WMT Paired Associates*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
PR	0.54768	5	0.113446	0.000543	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	58.4556		13.8628	4.217	0.000142
AMI ~ PR	0.5144		0.1477	3.484	0.001235
(Intercept)	34.5508		19.3083	1.789	0.081314
VMI ~ PR	0.7589		0.2057	3.690	0.000684
(Intercept)	46.8091		15.6955	2.982	0.00491
VWMI ~ PR	0.6336		0.1672	3.790	0.00051
(Intercept)	38.2889		15.6799	2.442	0.0192
IMI ~ PR	0.7311		0.1670	4.378	8.71e-05
(Intercept)	40.8229		15.5228	2.630	0.012163
DMI ~ PR	0.7129		0.1653	4.312	0.000107

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Free Recall score of the WMT. Free Recall data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted

of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 2.7719$, $p < 0.03273$), with an R^2 of 0.064543. Table 20 presents a regression model predicting WMS-IV index scores from WMT Free Recall.

Table 20: *Regression model predicting WMS-IV index scores from WMT Free Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
FR	0.71634	5	0.064543	0.03273	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	86.3778		6.8340	12.639	2.28e-15
AMI ~ FR	0.3105		0.1030	3.013	0.00453
(Intercept)	83.7589		10.1150	8.281	4.01e-10
VMI ~ FR	0.3343		0.1525	2.192	0.0344
(Intercept)	91.0308		8.4268	10.803	2.75e-13
VWMI ~ FR	0.2308		0.1271	1.817	0.077
(Intercept)	84.5619		8.4125	10.052	2.21e-12
IMI ~ FR	0.3396		0.1268	2.677	0.0108
(Intercept)	83.3092		8.1064	10.277	1.17e-12
DMI ~ FR	0.3717		0.1222	3.041	0.0042

* = Significant at 0.0025 level

A multivariate linear regression was calculated to predict performance on composite scores from the WMS-IV based on performance on the Long Delayed Free Recall score of the WMT. Long Delayed Free Recall score data was used as a predictor for WMS-IV composite score dependent variables. WMS-IV composite score data consisted of the Auditory Memory Index (AMI), Visual Memory Index (VMI), Visual Working Memory Index (VWMI), Immediate Memory Index (IMI), and the Delayed Memory Index (DMI). No statistically significant regression equation was found ($F(5,35) = 1.7933$, $p < 0.1398$), with an R^2 of 0.044591. Table 21 presents a regression model predicting WMS-IV index scores from WMT Long Delayed Free Recall.

Table 21: *Regression model predicting WMS-IV index scores from WMT Long Delayed Free Recall*

Variable	Wilks	DVs	Overall r-squared	Pr(>F)	Significant at 0.0025
LDFR	0.79605	5	0.044591	0.1398	–
	Coefficient		Standard Error	Test Statistic	p-value
(Intercept)	86.2472		7.0675	12.203	6.83e-15
AMI ~ LDFR	0.3070		0.1048	2.929	0.00566
(Intercept)	88.0971		10.6573	8.266	4.19e-10
VMI ~ LDFR	0.2628		0.1581	1.662	0.105
(Intercept)	94.581		8.833	10.71	3.57e-13
VWMI ~ LDFR	0.173		0.131	1.32	0.194
(Intercept)	84.9785		8.7280	9.736	5.42e-12
IMI ~ LDFR	0.3274		0.1295	2.529	0.0156

(Intercept)	87.1497	8.6634	10.060	2.16e-12
DMI ~ LDFR	0.3070	0.1285	2.389	0.0218

* = Significant at 0.0025 level

Means of Variables

R_5 Does the examination of the means of all variables in the obtained sample look comparable to the means found in the literature?

The mean, standard deviation, and range for each WMS-IV, TOMM, and WMT variable is presented in Table 22 and Table 23.

Table 22: *Mean, Standard Deviation, and Range Statistics for the WMS-IV*

Variable	Mean	SD	Min	Max
WMS-IV Subtests				
Logical Memory I	11.58	2.39	3	15
Logical Memory II	10.98	2.20	5	15
Verbal Paired Associates I	10.71	2.19	3	14
Verbal Paired Associates II	11.16	1.82	4	13
Designs I	11.24	2.81	5	19
Designs II	11.16	2.70	6	17
Visual Reproduction I	10.36	2.76	4	14
Visual Reproduction II	10.64	2.85	6	17
Spatial Addition	10.57	2.67	1	14
Symbol Span	11.50	2.10	7	16
Designs I Content	11.36	2.25	6	17
Designs I Spatial	10.66	3.25	5	17
Designs II Content	11.34	2.51	3	18

Designs II Spatial	10.45	3.05	6	17
WMS-IV Composites				
Auditory Memory Index	106.30	9.92	78	126
Visual Memory Index	105.09	13.93	73	132
Visual Working Memory Index	106.21	11.42	63	126
Immediate Memory Index	106.78	11.94	72	130
Delayed Memory Index	106.93	11.93	72	128

Note. Subtests have a mean of 10 and SD of 3. Composites have a mean of 100 and SD of 15.

Table 23: *Mean, Standard Deviation, and Range Statistics for the TOMM and WMT*

Variable	Mean	SD	Min	Max
WMT Scores				
Immediate Recognition	98.07	4.07	78	100
Delayed Recognition	98.95	2.91	83	100
Consistency	97.82	4.01	80	100
Multiple Choice	93.18	9.65	65	100
Paired Associates	93.07	10.24	60	100
Free Recall	64.59	14.22	35	100
Long Delay Free Recall	65.66	13.77	38	95
TOMM Scores				
Trial I	48.98	1.31		
Trial II	49.98	0.16		
Sum of Trial I and II	98.95	1.30		

Note. WMT scores reported as percentage correct. TOMM scores reported as raw scores.

Correlation Results

R_6 How much correlation is present between TOMM and WMT scores?

Correlations between TOMM and WMT scores were calculated. Two methodologies for computing correlations were performed. Pearson correlations were obtained as this is a commonly used method for assessing the strength of a linear association between two variables. Spearman correlations were obtained in addition to Pearson correlations as Spearman correlations are rank-order correlations that are considered by some researchers to be a better methodology for analyzing data that does not adhere to a normal distribution (Merten, Bossink, & Schmand, 2007). In the present study, the obtained TOMM and WMT scores do not appear to be normally distributed. Pearson correlations and Spearman correlations are presented in Table 24 and Table 25.

Table 24: *Pearson correlations between the TOMM and WMT*

Variable	1	2	3	4	5	6	7	8	9	10
WMT Scores										
1. Immediate Recognition	---									
2. Delayed Recognition	0.77†	---								
3. Consistency	0.97†	0.77†	---							
4. Multiple Choice	0.56†	0.63†	0.61†	---						
5. Paired Associates	0.63†	0.67†	0.67†	0.92†	---					
6. Free Recall	0.34*	0.45†	0.40†	0.54†	0.56†	---				
7. Long Delay Free Recall	0.19	0.28	0.26	0.47†	0.44†	0.80†	---			
TOMM Scores										
8. Trial I	0.17	0.14	0.15	0.02	0.03	-0.13	-0.14	---		
9. Trial II	-0.08	-0.06	-0.09	-0.12	-0.11	-0.24	-0.29	-0.12	---	
10. Sum of Trial I and II	0.17	0.13	0.14	0.01	0.01	-0.16	-0.18	0.99†	-0.01	---

† = Significant at 0.01 * = Significant at 0.05

Table 25: *Spearman correlations between the TOMM and WMT*

Variable	1	2	3	4	5	6	7	8	9	10
WMT Scores										
1. Immediate Recognition	---									
2. Delayed Recognition	0.50†	---								
3. Consistency	0.90†	0.60†	---							
4. Multiple Choice	0.46†	0.47†	0.56†	---						
5. Paired Associates	0.46†	0.44†	0.57†	0.84†	---					
6. Free Recall	0.17	0.45†	0.32*	0.58†	0.53†	---				
7. Long Delay Free Recall	0.17	0.39†	0.29	0.61†	0.55†	0.78†	---			
TOMM Scores										
8. Trial I	0.25	0.09	0.21	0.12	0.11	0.01	-0.08	---		
9. Trial II	-0.13	-0.09	-0.13	-0.15	-0.16	-0.26	-0.24	-0.16	---	
10. Sum of Trial I and II	0.23	0.07	0.19	0.09	0.07	-0.05	-0.13	0.98†	0.06	---

† = Significant at 0.01 * = Significant at 0.05

R_7 How much correlation is present between WMS-IV, TOMM, and WMT scores?

Pearson correlations were used to explore the relationship between WMS-IV and TOMM as well as between the WMS-IV and WMT. Results from WMS-IV, TOMM, and WMT correlations are presented in Table 26 and Table 27.

Table 26: *Pearson correlations between the WMS-IV and TOMM*

Variable	1	2	3	4	5	6	7	8
WMS-IV Composites								
1. Auditory Memory	---							
2. Visual Memory	0.47†	---						
3. Visual Working Memory	0.19	0.50†	---					
4. Immediate Memory	0.80†	0.85†	0.47†	---				
5. Delayed Memory	0.75†	0.88†	0.40†	0.85†	---			
TOMM Scores								
6. Trial I	0.01	0.05	0.07	0.05	0.05	---		
7. Trial II	0.07	-0.30	-0.12	-0.16	-0.14	-0.12	---	
8. Sum of Trial I and II	0.02	0.01	0.05	0.03	0.03	0.99†	-0.01	---

† = Significant at 0.01 * = Significant at 0.05

Table 27: *Pearson correlations between the WMS-IV Composites and the WMT*

Variable	1	2	3	4	5	6	7	8	9	10	11	12
WMS-IV Composites												
1. AMI	---											
2. VMI	0.47†	---										
3. VWMI	0.19	0.50†	---									
4. IMI	0.80†	0.85†	0.47†	---								
5. DMI	0.75†	0.88†	0.40†	0.85†	---							
WMT Scores												
6. IR	0.40†	0.41†	0.53†	0.44†	0.46†	---						
7. DR	0.40†	0.35*	0.51†	0.41†	0.42†	0.75†	---					
8. CNS	0.41†	0.40†	0.46†	0.46†	0.45†	0.94†	0.82†	---				
9. MC	0.49†	0.44†	0.46†	0.54†	0.51†	0.58†	0.66†	0.64†	---			
10. PR	0.51†	0.51†	0.53†	0.58†	0.58†	0.65†	0.69†	0.69†	0.92†	---		
11. FR	0.46†	0.34*	0.29	0.43†	0.46†	0.34*	0.44†	0.40†	0.54†	0.56†	---	
12. LDFR	0.44†	0.27	0.22	0.39†	0.37*	0.19	0.28	0.26	0.47†	0.44†	0.80†	---

† = Significant at 0.01 * = Significant at 0.05

Chapter V

DISCUSSION

Discussion

This chapter is divided into 4 sections: summary of the study, discussion of the results, limitations and delimitations of the study, and directions for future research.

Purpose of the study. The purpose of this study was to explore the relationship between a well-validated measure of memory and commonly used symptom validity tests in neuropsychology. Specifically, this study examined how performance on the *Test of Memory Malingering* (TOMM; Tombaugh, 1996) and *Word Memory Test* (WMT; Green, 2003) was related to performance on the *Wechsler Memory Scale, Fourth Edition* (WMS-IV; Wechsler, 2009). This study also explored the relationship within and between the TOMM and the WMT. Data for this study came from 46 students attending a psychology class at a Midwestern university. These students were instructed to provide full effort for the entirety of their session and were offered research participation credit regardless of completion of all tasks. Student demographics and performance were largely consistent with what would be expected for a Midwestern university sample. Students were interviewed for demographic information, administered measures of memory and symptom validity as well as other related assessments for a larger study. Students typically completed all tasks in nearly 4 hours and were not held past 4 hours.

A total of 20 multivariate regression analyses were used to address 4 research questions examining the relationship between the WMS-IV, TOMM or WMT. An alpha of 0.0025 was selected due to necessity of using a Bonferroni correction. A Bonferroni correction was selected due to it being regarded as a conservative method for avoiding Type I errors (Dunn, 1961). Normality and linearity of data were examined by Q-Q Plots and Mardia tests of multivariate skew and kurtosis (Mardia, 1970).

Spearman and Pearson correlations were used to examine the relationship between the TOMM and WMT. A Spearman correlation was conducted in addition to a Pearson correlation as Spearman correlations are rank-order correlations that are considered by some researchers to be better suited to data that does not align with a normal distribution (Merten, Bossink, & Schmand, 2007). It is typical for adults who take symptom validity tests who are not feigning impairment to provide restricted ranges of data on symptom validity indicators, such as the TOMM (Hill, Laurie, Kennedy, & Malamut, 2003; Morgan & Sweet, 2008; Brooks, Sherman, & Krol, 2012). This pattern of restricted ranges of data on the TOMM and WMT was observed in the present study.

Means, standard deviations, and frequencies of demographic characteristics were computed. All calculations and statistical analyses were performed in R i386 3.3.2 (R Core Team, 2013) with relevant add-on packages.

Discussion of the results. Upon examination of the data obtained in this study and examination of relevant literature (Hill, Laurie, Kennedy, & Malamut, 2003; Morgan & Sweet, 2008; Brooks, Sherman, & Krol, 2012), it was clear that patterns of data could hinder usage of common parametric statistics. The symptom validity tests generally showed a restricted range in performance. For example, on TOMM Trial 2 the mean was 49.98 and the standard deviation was 0.16. Given the symptom validity tests should be easily completed by the vast majority of participants who are not feigning impairment, the restricted range of performance was expected and the reduced variability makes finding a relationship between effort and memory measures less likely. It is important to note

that the lack of a significant relationship is useful for clinicians as these tests should be measuring unrelated constructs. A discussion of the statistical analyses for each research question is presented below.

WMS-IV subtest variance accounted for by the TOMM. A multivariate regression was used to explore how well TOMM Trial 1 and Trial 2 could account for variance in WMS-IV subtests. In the present study and in many clinical applications, data on TOMM Trial 1 and Trial 2 are available, so a summation of both trials was also calculated in an effort to increase the variability and possibly other statistical properties (e.g. reliability). This effort did not make meaningful improvements in variability for the present study. The unsuccessful attempts to mathematically increase chances of a relationship between the TOMM and WMS-IV reinforce the idea that the TOMM and WMS-IV are measuring unrelated constructs.

The scores obtained from the TOMM were used as predictors for performance on WMS-IV subtests during these analyses. For TOMM Trial 1, no statistically significant regression equation was found ($F(16,21) = 1.2745$, $p < 0.2959$), with an R^2 of 0.041523. For TOMM Trial 2, no statistically significant regression equation was found ($F(16,21) = 0.89061$, $p < 0.5877$), with an R^2 of 0.031853. For the summation of TOMM Trial 1 and Trial 2, no statistically significant regression equation was found ($F(16,21) = 1.3988$, $p < 0.2326$), with an R^2 of 0.04433. Since 20 separate regressions were conducted, an alpha level of 0.0025 instead 0.05 was used. Adjusting the alpha level through a Bonferroni correction is a conservative way to reduce chances for Type I errors.

These results have several implications for researchers and clinicians. The results suggest that the TOMM will likely not be predictive of performance on measures of memory, especially in a population resembling the sample for this study. This suggests that for individuals passing symptom validity indicators, the results of administered memory measures will very likely be autonomous from symptom validity indicator performance. From a perspective of construct validity, the TOMM being

independent from the subtests on the WMS-IV is reassuring as these tests were designed to measure different constructs. The present study; however, does not show if the TOMM and WMS-IV would be related in samples providing suboptimal performance and the results may be different from clinical populations or individuals of different ages.

WMS-IV subtest variance accounted for by the WMT. A multivariate regression was used to explore if WMT scores account for variance in WMS-IV subtests. A total of 7 WMT scores were used to predict WMS-IV subtest performance. No statistically significant regression equation was found for any of the WMT scores: WMT Immediate Recall ($F(16,24) = 1.4183$, $p < 0.2137$, $R^2 = 0.040743$), WMT Delayed Recall ($F(16,24) = 1.9006$, $p < 0.07522$, $R^2 = 0.049869$), WMT Consistency ($F(16,24) = 1.2062$, $p < 0.3305$, $R^2 = 0.03621$), WMT Multiple Choice ($F(16,24) = 2.2879$, $p < 0.03247$, $R^2 = 0.056252$), WMT Paired Associates ($F(16,24) = 2.0906$, $p < 0.04972$, $R^2 = 0.053092$), WMT Free Recall ($F(16,24) = 1.4305$, $p < 0.2083$, $R^2 = 0.040994$), and WMT Long Delayed Free Recall ($F(16,24) = 0.86505$, $p < 0.6108$, $R^2 = 0.028057$).

While WMT scores were not predictive of WMS-IV scores, it was apparent that some of the WMT scores designed to measure memory (e.g. Paired Associates and Multiple Choice) rather than effort would have been significant without the application of a Bonferroni correction. Statistical significance would have been a reasonable finding given task similarity; however, the sheer number of analyses conducted raises the probability of Type I errors and changes the threshold for concluding statistical significance.

The implications of the findings of the predictive value of WMT scores for WMS-IV performance are similar to findings for the relationship between the TOMM and WMS-IV. The WMT and WMS-IV are designed to measure mostly different constructs. The autonomy of the WMT symptom validity measures from the WMS-IV reassures the clinician and researcher that the performance on a symptom validity indicator is not typically affected by participants' genuine level of

functioning. As noted in the discussion of the previous research question, the current study does not address if the relationship between the WMT and WMS-IV would change with a sample providing suboptimal performance, different ages, or clinical populations.

WMS-IV composite variance accounted for by the TOMM. A multivariate regression was used to explore how well each TOMM Trial 1 could account for variance in WMS-IV composite scores and how well TOMM Trial 2 could account for variance in WMS-IV composite scores. A summation of TOMM Trial 1 and Trial 2 was utilized for this analysis for the reasons stated in the first research question. An exploration of how well the TOMM accounts for WMS-IV composite score variance was conducted in addition to exploring the relationship between the TOMM and WMS-IV subtests as composite scores have different statistical properties than the individual subtests from which they are comprised, such as reliability, and other attributes desirable for clinicians and researchers.

The scores obtained from the TOMM were used as predictors for performance on WMS-IV composites during these analyses. For TOMM Trial 1, no statistically significant regression equation was found ($F(5,32) = 0.058231$, $p < 0.9976$), with an R^2 of 0.001811. For TOMM Trial 2, no statistically significant regression equation was found ($F(5,32) = 1.338$, $p < 0.2736$), with an R^2 of 0.037257. For the summation of TOMM Trial 1 and Trial 2, no statistically significant regression equation was found ($F(5,32) = 0.047738$, $p < 0.9985$), with an R^2 of 0.001484.

The relationship between the TOMM and WMS-IV composites appear to be similar or even weaker than the relationship between the TOMM and WMS-IV subtests. Given increased reliability of WMS-IV composites compared to WMS-IV subtests, these findings increase confidence that the TOMM and WMS-IV were not related in the present sample.

WMS-IV composite variance accounted for by the WMT. A multivariate regression was used to explore how well each WMT score could account for variance in the WMS-IV composite scores. As with the exploration of the WMS-IV composite and TOMM relationship, the relationship between the

WMT and WMS-IV composite scores would be expected to be more reliable than the relationship between the WMT and WMS-IV subtest scores.

The scores obtained from the WMT were used as predictors for performance on WMS-IV composites during these analyses. For the WMT Immediate Recall score, no statistically significant regression equation was found ($F(5,35) = 4.4651$, $p < 0.002988$), with an R^2 of 0.093967. For the WMT Delayed Recall score, no statistically significant regression equation was found ($F(5,35) = 4.0836$, $p < 0.005028$), with an R^2 of 0.087812. For the WMT Consistency score, no statistically significant regression equation was found ($F(5,35) = 3.2457$, $p < 0.01641$), with an R^2 of 0.07336. For the WMT Multiple Choice score, no statistically significant regression equation was found ($F(5,35) = 3.9367$, $p < 0.006162$), with an R^2 of 0.085376. For the WMT Paired Associates score, a statistically significant regression equation was found ($F(5,35) = 5.7812$, $p < 0.000543$), with an R^2 of 0.113446. For WMT Free Recall score, no statistically significant regression equation was found ($F(5,35) = 2.7719$, $p < 0.03273$), with an R^2 of 0.064543. For the WMT Long Delayed Free Recall, no statistically significant regression equation was found ($F(5,35) = 1.7933$, $p < 0.1398$), with an R^2 of 0.044591.

In the exploration of the WMT and WMS-IV composite scores, the WMT Paired Associates test predicted about 11% of variance in WMS-IV composite scores. With most WMT scores being unrelated to WMS-IV performance, it may be most conservative to conclude that the WMT does not serve as a good predictor of WMS-IV performance in this population.

Examination of the means of all variables and comparison to means found in the literature. When previous data on measures exists, it is important and interesting to see how data from a sample of participants compares to data reported from other researchers. In general, the mean scores of the sample fell close to available normative means and thus was considered largely representative of what would be expected from this university sample. Normative data is readily available in the literature for

the TOMM, WMT, and WMS-IV (e.g. Tombaugh, 1997; Rienstra, Spaan, & Schmand, 2009, Wechsler, 2009).

Correlation between TOMM and WMT scores. Pearson and Spearman correlations were used to explore the strength of correlations between TOMM and WMT scores. Spearman correlations were conducted in addition to Pearson correlations as these correlations are rank-order correlations and may be more resilient to non-normal data distributions such as what is found in TOMM and WMT distributions (Merten, Bossink, & Schmand, 2007).

Pearson and Spearman correlations were generally in agreement; however, the Spearman correlations were often lower. The correlations generally adhered to expected patterns. Immediate Recognition, Delayed Recognition, and Consistency on the WMT had strong correlations with one another based upon Cohen's (1988) guidelines. Multiple Choice, Paired Associates, Free Recall, and Long Delayed Free Recall also generally had strong correlations with one another. Correlations between the first three WMT scores and the other WMT scores generally had moderate correlations. This is not surprising as these tests are similar yet are intended to measure different constructs.

The TOMM scores had nearly perfect correlations or what appeared to be weak inverse correlations. All TOMM scores are intended to measure the same construct, so this pattern of data may initially appear surprising. The correlation findings demonstrate the importance of examining data (e.g. viewing scatterplots) and reviewing assumptions prior to using correlation analysis and prior to making any final conclusions. Correlational data can easily hide non-linear data patterns, subgroups present in the data, and potential data abnormalities (Bewick, Cheek, & Ball, 2003).

Correlation between WMS-IV composite scores, TOMM scores, and WMT scores. Pearson correlations were conducted to explore if there was a relationship between WMS-IV composite scores and scores from the TOMM and WMT. Specifically, a relationship between visual memory tests and the TOMM as well as a relationship between verbal memory tests and the WMT was examined.

All correlations between the TOMM and the WMS-IV composites were small or weak based upon Cohen's (1988) guidelines with the exception of a moderate inverse correlation between TOMM Trial 2 and the Visual Memory Index of the WMS-IV. This suggests that performance on the TOMM does not have a relationship with visual or verbal composites of the WMS-IV for this sample. The sole moderate inverse correlation was interesting, but was attributed to being a product of the non-normal TOMM data used in the study. If this relationship was a genuine relationship, it would suggest that a higher performance on the TOMM Trial 2, specifically, would be related to a lower performance on the visual memory subtests of the WMS-IV.

All correlations between the WMT and WMS-IV composites were in the moderate to large range with two exceptions barely being below the moderate range. Patterns of the WMT being more strongly related to any specific WMS-IV composites were not found. As with the exploration of the relationship between the TOMM and WMS-IV, the data suggests that the WMT is not more strongly related to WMS-IV tasks that appear similar in presentation than other WMS-IV tasks.

What is perhaps most interesting from these analyses is that correlations were present between WMT and WMS-IV composites. This suggests that a higher performance on the WMT is related to a higher performance on WMS-IV composites. In some cases, the WMT symptom validity scores are more highly correlated with the WMS-IV than WMT memory scores. For example, the Visual Working Memory Index from the WMS-IV and the Immediate Recall score from the WMT have a correlation of 0.53 whereas the Visual Working Memory Index and Long Delayed Free Recall score from the WMT have a correlation of 0.22. Upon review of the WMT and WMS-IV relationship explored through regression, it appears that the regression would have detected significant relationships; however, application of the Bonferroni correction leads to more stringent criteria and determinations of nonsignificance. Stringent criteria, such as a straightforward Bonferroni correction can be applied and provide effective protection from Type I errors; however, it remains debatable if a

Bonferroni correction would be too strict. Significant literature has been produced discussing multiple comparison problems. For example, Lindquist and Mejia (2015), explore using appropriate corrections during multiple comparisons. Lindquist and Mejia suggest that in cases where comparisons are not completely independent, a Bonferroni correction, can be too strict. In the present study, many of the comparisons are clearly not independent. One example is in the composition of WMS-IV composite scores. Some WMS-IV composite scores are based on overlapping subtests. Even when there is not clear overlap, one can intuitively conclude that individually measured domains of memory would not be independent.

Limitations

The current study had a number of limitations that could impact generalizability. For example, the sample was derived from a nonclinical Midwestern university population. It is plausible that clinical populations may exhibit greater variability in memory performance and symptom validity performance. Since limited variability was present in the present study, a greater variability of performance in a clinical population may elucidate stronger relationships between memory and symptom validity performance.

The present study did not utilize experimental research methodology. An experimental or quasi-experimental study with participants assigned to separate conditions would allow for greater confidence in results and greater confidence in explaining obtained data. For example, studies such as those performed by Suhr and colleagues (e.g. Suhr & Gunstad, 2000; Suhr, Gunstad, Greub, & Barrash, 2004) assigned participants to groups instructed to provide full effort, to feign memory impairment, or to feign memory impairment after coaching. Performance differences in experimental groups can often be attributed to the condition or conditions manipulated in the study. The present study did not allow for these type of between-group or within-group comparisons; all participants were given instructions to provide full effort.

The present study had a modest sample size of 46 participants. Ideally, regression analyses should have hundreds, if not thousands, of participants in many cases (Knofczynski & Mundfrom, 2008). Using additional groups for experimental conditions would have required an even larger number of participants for confident statistical interpretation. Recruiting and administering assessments to additional participants is potentially burdensome for small research teams. Indeed the assessment administration time alone for 46 participants required approximately 184 hours of time.

An additional important consideration for research is diversity of the sample. The sample in the present study had a nearly equal representation of men (46%) and women (54%), but other sample demographics stray farther from the ideal sample. The present sample consisted solely of undergraduate students at one Midwestern university which limits generalizability to similar groups. To increase generalizability additional samples would consist of participants with varying education levels, ages, and geographic diversity. Including participants from more settings across a wider geographic region may make the ethnic diversity of the sample match the general population. The present study, for example, included two participants (4%) who identified as Hispanic or Latino. As of 2016, about 17.8% of the United States population identified as Hispanic or Latino (U.S. Census Bureau, 2016). Further consideration of the diversity of functioning in the sample is also important for generalizability. For example, the present sample had 9% of participants self-reporting ADHD. Meta-analyses in the literature, such as the one presented by Simon, Czobor, Bálint, Mészáros, and Bitter (2009), suggests that adult ADHD prevalence is 2.5% with a range of 1 to 7.3%. The present sample is somewhat larger than this; however, the small sample size, reasonably strong representation of males, and young adulthood ages of the present sample make the present findings reasonable.

Delimitations

The present study has a number of strengths that increase applicability to research and clinical settings. The current study utilizes memory and symptom validity measures that have already been

carefully designed, studied, and used in research and clinical applications. The WMS-IV is one of the most popular measures of memory used in clinical settings (Rabin, Paolillo, & Barr, 2016). The TOMM is regarded as the most widely used and studied symptom validity indicator (Jelicic, Ceunen, Peters, & Merckelbach, 2011). The WMT is a symptom validity indicator that has numerous validation studies on clinical samples (Green, Lees-Haley, & Allen, 2003).

The measures utilized in the present study are appropriate for the sample used. The literature suggests that the symptom validity indicators used in the present study are not significantly impacted by ability levels. For example, the WMT has been shown to be insensitive to all but most extreme impairments in learning and memory (Flaro, Green, & Robertson, 2007). Although the present university student sample had memory functioning that was modestly higher than the mean (e.g. WMS-IV Auditory Memory Index mean = 106.30), it is likely that symptom validity performance for a nonclinical population with similar demographics instructed to provide full effort would yield similar data.

The present study utilized a university sample. Although this sample differs from a clinical sample, it also presents some of its own strengths. For this particular sample, there is no known incentive to malinger. This fact means that the findings of this study is likely applicable to other research and clinical populations with no known incentive to malinger. The present university sample also has a fairly consistent level of education (mean years of education = 13.00, standard deviation = 1.30) and known correlates of ability such as SAT scores. The findings of this study can be regarded as being generalizable to other non-clinical adult populations with average to above average educational attainment and abilities. The literature suggests that the symptom validity tests used in this study are resilient to educational and ability level, so findings of this study may be applicable to populations with more variance in these domains. Researchers, such as Banerjee and Chaudhury (2010), caution that results from samples should only be generalized to populations that sample was taken from; however, it

has been established these measures as performing similarly across age and ability levels. Green and Flaro (2003), for example, have demonstrated that even children can pass the WMT if they have a 3rd grade reading level. Hill, Laurie, Kennedy, and Malamut (2003) have demonstrated that the TOMM can be passed with by individuals with temporal lobe dysfunction.

The present study consisted assessments that took approximately four hours to administer. Initially, a lengthy assessment can present as a weakness as participants and examiners could experience fatigue effects; however, the length of the assessment adds to the ecological validity of the study. Multiple assessments are often given during clinical neuropsychological examinations which can result in sessions spanning several hours. Symptom validity tests were interspersed in the present study which was consistent with how these tests are designed to be administered. In addition, test batteries were administered in the same order for each participant.

Future Directions

The present research addresses some questions related to the relationships between symptom validity measures and memory measures; however, there are a number of avenues available for future research. Some studies (e.g. Suhr & Gunstad, 2000; Suhr, Gunstad, Greub, & Barrash, 2004) have experimental conditions in which some participants are instructed to malinger, some participants are instructed to malinger and given coaching, and some participants are instructed to provide full effort. Adding experimental conditions to a study similar to the present study could add a number of significant strengths. Researchers could determine if symptom validity indicators hold the same relationship with memory measures in a variety of conditions. With some participants providing suboptimal effort, it would be likely that some participants would fail symptom validity tests and provide greater variability in scores for both symptom validity tests and memory measures. The greater

variability in scores could potentially lead to more normal distributions in data and the ability to apply different statistical tests.

The addition of experimental conditions in which participants are instructed to malinger could also build confidence in the application of research findings to clinical populations. In some referred populations, participants might provide suboptimal effort and possibly fail symptom validity indicators. Referred populations may be seeking external incentives via successful malingering whereas experimental research samples may not have strong incentives to malinger.

Future studies could also include samples from clinical populations and not just university populations. Clinical samples used in future research could include participants with diversity in age, ability, referral concerns, and incentives to malinger. This would result in larger sample sizes as data could be pulled from cases spanning years. Of course, this would likely limit opportunities for experimental manipulations and reasonable certainties of incentives provided and effort given.

Future studies may benefit from usage of other types of statistics and corrections that are available or may become available. A larger sample size with more variance in performance could open more doors for statistical methods and applications of mathematical data transformations could allow for data to better meet assumptions for statistical analysis. Future studies could also include explorations of different memory and symptom validity indicators.

The present study looked at predicting memory performance based upon symptom validity test performance. It may be more useful to clinicians to reverse predictive directionality. Clinicians may routinely use memory or related cognitive assessments and could ponder whether or not symptom validity tests should be administered. Indeed, some embedded measures predicting symptom validity based on memory or cognitive assessment performance already exist. Simple embedded measures such as Reliable Digit Span (RDS) have been found to be strongly related to symptom validity tests and more useful than other available embedded measures (Miele, Gunner, Lynch, & McCaffrey,

2012).Further improvement of embedded measures could shorten overall assessment time and expenditures and provide confidence in validity of obtained data.

Summary of Implications for Clinical Practice and Future Research

The present study suggests that for individuals providing adequate effort during assessments, it is unlikely that memory performance can be predicted based upon symptom validity performance. Symptom validity tests, such as those employed in this study, are typically designed to appear as difficult memory tests but are easily passable in reality. In the present sample and likely in many populations, participants providing adequate effort during assessments obtain passing scores on symptom validity tests. Memory assessments often allow for precise measurement of a wide variety of performances in clinical and healthy research samples. A variety of scores cannot typically be predicted from scores that do not vary. Metaphorically, intensity of light in every room of a university cannot be predicted based upon knowledge that every light switch is completely turned on.

For clinical and research settings, the data suggests that symptom validity test performance is not predictive of performance on memory tests. It may, however, be possible to use embedded symptom validity measures in memory tests to predict performance. Again revisiting the metaphor above, it may be possible in some situations to predict if a light switch is turned off based upon the intensity of light measured in a room. For the foreseeable future, one of the best ways to measure symptom validity during an examination is to use the TOMM or WMT or another measure specifically designed for this purpose. One of the best ways to measure memory performance is also to administer the WMS-IV or another measure specifically designed for this purpose.

Data from the present study suggests that scores from one symptom validity indicator are not effectively comparable with scores from another symptom validity indicator via correlation for the current sample under study. Correlation and a number of statistical methods do not work as expected when assumptions, such as normality of data, are not well satisfied. In clinical settings, in particular, it

may be most important to simply know if adequate effort is given (binary data) rather than a very specific measurement of symptom validity given at that point in time (continuous data). Comparisons between symptom validity indicators, should therefore, focus on agreement rates (binary data) when comparisons of continuous data from these measures is not possible for statistical or theoretical reasons.

Summary of the Study

The sample obtained for this study had fairly unremarkable differences in performance compared to what would be expected based upon the norms available for memory and symptom validity measures utilized for this study. Given the sample was college students rather than a sample from the general population, the small positive skew found on memory measure performance was not particularly noteworthy. Effort test results had a restricted range of data. While problematic for statistical analysis, the restricted range of data for symptom validity tests suggests valid administration of the research protocol.

The regression analyses conducted during this research showed that a number of memory and symptom validity variables were related; however, the sheer number of analyses conducted increased likelihood of false positive results. One conservative option for correcting for false positives, is the Bonferroni correction. After application of this simple correction, there were still some significant results such as the relationship between all WMT scores and the WMS-IV composite scores. While this is interesting, careful examination of data patterns suggested that assumptions for usage of regression were violated. With this in mind, it is likely that false positive results are still possible. Logarithmic data transformations may prove successful in improving data linearity; however, the extremely restricted range of symptom validity test performance likely cannot be corrected mathematically and therefore trumps all other concerns.

Correlational data between symptom validity indicators was also problematic due to non-normal data patterns. Similar to the findings of the regression analyses, the restricted range of performance on symptom validity indicators presented difficulties for both types of attempted correlational analyses. Well-known methods of statistical analysis are not effective when assumptions for the data are violated. Indeed the best chance of getting usable statistical information is careful research design and consideration of the research questions.

References

Aguerrevere, L. E., Greve, K. W., Bianchini, K. J & Ord, J. S. (2011) Classification accuracy of the Millon Clinical Multiaxial Inventory-III modifier indices in the detection of malingering in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 33*(5), 497-504. doi:10.1080/13803395.2010.535503

Allen, M. D., Bigler, E. D., Larsen, J., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2007). Functional neuroimaging evidence for high cognitive effort on the Word Memory Test in the absence of external incentives. *Brain Injury, 21*(13/14), 1425-1428. doi:10.1080/02699050701769819

Allen, L. M., Conder, R. L., Green, P., & Cox, D. R. (1997). *CARB 97 manual for the Computerized Assessment of Response Bias*. Durham, NC: CogniSyst.

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders* (1st ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (1968). *Diagnostic and statistical manual of mental disorders* (2nd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Arlington, VA: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., text rev.). Arlington, VA: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Arlington, VA: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Arlington, VA: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.

American Psychological Association. (2010). *Ethical principles of psychologists and code of conduct.* Retrieved from http://apa.org/ethics/code/index.aspx

Armistead-Jehle, P., & Gervais, R. O. (2011). Sensitivity of the Test of Memory Malingering and the Nonverbal Medical Symptom Validity Test: A Replication Study. *Applied Neuropsychology, 18*(4), 284-290. doi: 10.1080/09084282.2011.595455

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. Industrial Psychiatry Journal, 19(1), 60–65. http://doi.org/10.4103/0972-6748.77642

Batt, K., Shores, E., & Chekaluk, E. (2008). The effect of distraction on the Word Memory Test and Test of Memory Malingering performance in patients with a severe brain injury. *Journal of The International Neuropsychological Society, 14*(6), 1074-1080. doi: 10.1017/S135561770808137X

Bentler, P. M., & Yuan, K. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research, 34*(2), 181-197. doi:10.1207/S15327906Mb340203

Bewick, V., Cheek, L., & Ball, J. (2003). Statistics review 7: Correlation and regression. Critical Care, 7(6), 451–459.

Brooks B. L., Sherman E. M., Krol A. L. (2012). Utility of TOMM Trial 1 as an indicator of effort in children and adolescents. *Archives of Clinical Neuropsychology,27*(1):23-29. doi: 10.1093/arclin/acr086

Bush, S. S., Ruff, R. M., Tröster, A. I., Barth, J. T., Koffler, S. P., Pliskin, N. H., Reynolds, C.R., & Silver, C. H. (2005). Symptom validity assessment: Practice issues and medical necessity: NAN Policy & Planning Committee. *Archives of Clinical Neuropsychology, 20*(4), 419-426.

Butcher, J. N., Graham, J. R., Ben-Porath, Y. S., Tellegen, A., Dahlstrom, W. G., & Kaemmer, B. (2001). *Minnesota Multiphasic Personality Inventory-2: Manual for administration and scoring* (revised ed.). Minneapolis, MN: University of Minnesota Press.

Carone, D. A., & Bush, S. S. (2013). Introduction: Historical perspectives on mild traumatic brain injury, symptom validity assessment, and malingering. In D. A. Carone, S. S. Bush (Eds.), *Mild traumatic brain injury: Symptom validity assessment and malingering* (pp. 1-29). New York, NY US: Springer Publishing Co.

Clauss-Ehlers, C. (2008). Encyclopedia of cross-cultural school psychology. New York London: Springer.

Cohen, M. J. (1997). *Children's Memory Scale*. San Antonio, TX: The Psychological Corporation.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Toronto: Holt, RineHart, and Winston, Inc.

Delis, D.C., Kramer, J.H., Kaplan, E., & Ober, B.A. (1987). The California Verbal Learning Test. New York: The Psychological Corporation.

Denning, J. H. (2012). The efficiency and accuracy of The Test of Memory Malingering Trial 1, errors on the first 10 Items of The Test of Memory Malingering, and five embedded measures in predicting invalid test performance. *Archives of Clinical Neuropsychology, 27*(4), 417-432. doi:10.1093/arclin/acs044

Drozdick, L., Holdnack, J. & Hilsabeck, R. (2011). *Essentials of WMS-IV assessment*. Hoboken, N.J: Wiley.

Faust, D. & Ziskin, J. (2011). Coping with psychiatric and psychological testimony: Based on the original work by Jay Ziskin. Oxford New York: Oxford University Press.

Fishbain, D., Cutler, R., Rosomoff, H., & Rosomoff, R. (2004). Is there a relationship between nonorganic physical findings (Waddell signs) and secondary gain/malingering?. *The Clinical Journal of Pain, 20*(6), 399-408. doi: 10.1097/00002508-200411000-00004

Flaro, L., Green, P., & Robertson, E. (2007). Word Memory Test failure 23 times higher in mild brain injury than in parents seeking custody: The power of external incentives. *Brain Injury, 21*(4), 373-383. doi: 10.1080/02699050701311133

Frederick, R.I. & Speed, F.M. (2007). On the Interpretation of Below-Chance Responding in Forced-Choice Tests. *Assessment, 14*(1), 3-11. doi: 10.1177/1073191106292009

functional magnetic resonance imaging. (2007). In The American Heritage Medical Dictionary. Retrieved from http://www.credoreference.com/entry/hmmedicaldict/functional_magnetic_resonance_imaging

Gervais, R. O., Rohling, M. L., Green, P., & Ford, W. (2004). A comparison of WMT, CARB, and TOMM failure rates in non-head injury disability claimants. *Archives of Clinical Neuropsychology, 19*(4), 475-487. doi: 10.1016/j.acn.2003.05.001

Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2009). Word memory test performance in amnesic patients with hippocampal damage. *Neuropsychology, 23*(4), 529-534. doi:10.1037/a0015444

Green, P. (2003). *Green's Word Memory Test for Windows: User's manual*. Edmonton, Alberta, Canada: Green's Publishing.

Green, P., & Flaro, L. (2003). Word Memory Test Performance in Children. Child *Neuropsychology, 9*(3), 189-207. doi:10.1076/chin.9.3.189.16460

Green, P., Iverson, G.L., & Allen, L.M. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13,* (10), 813-819. doi:10.1080/026990599121205

Green, P., Lees-Haley, P. R., & Allen, L. (2002). The Word Memory Test and the validity of neuropsychological test scores. *Journal of Forensic Neuropsychology, 2*(3-4), 97-124. doi:10.1300/J151v02n03_05

Green, P., Montijo, J., & Brockhaus, R. (2011). High Specificity of the Word Memory Test and Medical Symptom Validity Test in Groups with Severe Verbal Memory Impairment. *Applied Neuropsychology, 18*(2), 86-94. doi:10.1080/09084282.2010.523389

Green, P., Rohling, M. L., Lees-Haley, P. R., & Allen III, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045-1060. doi:10.1080/02699050110088254

Greer, S., Chambliss, L., Mackler, L., & Huber, T. (2005). Clinical inquiries. What physical exam techniques are useful to detect malingering?. *The Journal of Family Practice*, *54*(8), 719-722.

Gorissen, M., Sanz de la Torre, J.C. & Schmand, B. (2003). Effort in the Neuropsychological evaluation of schizophrenia. *Journal of the International Neuropsychological Society, 9*, 512.

Gunner, J. H., Miele, A. S., Lynch, J. K., & McCaffrey, R. J. (2012). The Albany Consistency Index for the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, *27*(1), 1-9. doi:10.1093/arclin/acr089

Haggerty, K., Frazier, T., Busch, R., & Naugle, R. (2007). Relationships among Victoria Symptom Validity Test Indices and Personality Assessment Inventory Validity Scales in a large clinical sample. *The Clinical Neuropsychologist, 21*(6), 917-928. doi:10.1080/13854040600899724

Halligan, P., Bass, C. & Oakley, D. (2003). *Malingering and illness deception*. New York: Oxford University Press.

Hartman, D. E. (2002). The unexamined lie is a lie worth fibbing Neuropsychological malingering and the Word Memory Test. *Archives of Clinical Neuropsychology, 17*(7), 709. doi:10.1093/arclin/17.7.709

Hathaway, S. R., & McKinley, J. C. (1942). *Minnesota Multiphasic Personality Inventory* manual (Rev. ed.). New York: Psychological Corporation.

Heaton, R. K., Smith, H. H., Lehman, R. A., & Vogt, A. T. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology, 46*(5), 892-900. doi:10.1037/0022-006X.46.5.892

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American Academy of Clinical Neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *The Clinical Neuropsychologist, 23*(7), 1093-1129. doi:10.1080/13854040903155063

Henry, G. K., Heilbronner, R. L., Mittenberg, W., Enders, C., Stevens, A., & Dux, M. (2011). Noncredible Performance in Individuals with External Incentives: Empirical Derivation and Cross-Validation of the Psychosocial Distress Scale (PDS). *Applied Neuropsychology, 18*(1), 47-53. doi:10.1080/09084282.2010.523385

Hill, S. L., Laurie, R. M., Kennedy, C. H., & Malamut, B. L. (2003). The relationship between measures of declarative memory and the Test of Memory Malingering in patients with and without temporal lobe dysfunction. *Journal of Forensic Neuropsychology, 3*(3), 1. doi:10.1300/J151v03n03_01

Hoelzle, J. B., Nelson, N. W., & Smith, C. A. (2011). Comparison of Wechsler Memory Scale–Fourth Edition (WMS–IV) and Third Edition (WMS–III) dimensional structures: Improved ability to evaluate auditory and visual constructs. *Journal of Clinical and Experimental Neuropsychology, 33*(3), 283-291. doi:10.1080/13803395.2010.511603

Holtz, J. (2011). Applied clinical neuropsychology an introduction. New York: Springer Pub. Co.

Iverson, G. L. (2003). Detecting malingering in civil forensic evaluations. In A. M. Horton Jr. & L. C. Hartlage (Eds.), *Handbook of forensic neuropsychology* (pp. 137–177). New York: Springer Publishing Company.

Iverson, G., Green P. & Gervais, R. (1999). Using the Word Memory Test to detect biased

responding in head injury litigation. *Journal of Cognitive Rehabilitation, 17*(2), 4-8.

Jelicic, M., Ceunen, E., Peters, M. V., & Merckelbach, H. (2011). Detecting coached feigning using the test of Memory Malingering (TOMM) and the structured inventory of Malingered Symptomatology (SIMS). *Journal of Clinical Psychology, 67*(9), 850-855.

Ju, D., & Varney, N. (2000). Can head injury patients simulate malingering?. *Applied Neuropsychology, 7*(4), 201-207. doi:10.1207/S15324826AN0704_1

Kelley, T. L. (1927). Interpretation of educational measurements. Oxford England: World Book Co.

Knofczynski, G. T., & Mundfrom, D. (2008). Sample Sizes When Using Multiple Linear Regression for Prediction. *Educational & Psychological Measurement, 68*(3), 431-442.

Lande, R., & Williams, L. (2013). Prevalence and characteristics of military malingering. *Military Medicine, 178*(1), 50-54. doi:10.7205/MILMED-D-12-00138

Larrabee, G. J. (2003). Exaggerated MMPI-2 symptom report in personal injury litigants with malingered neurocognitive deficit. *Archives of Clinical Neuropsychology, 18*(6), 673. doi:10.1016/S0887-6177(02)00157-9

Larrabee, G. J. (2007). Introduction: Malingering, research designs, and base rates. In G. J. Larrabee (Ed.) , *Assessment of malingered neuropsychological deficits* (pp. 3-13). New York, NY US: Oxford University Press.

Larsen, J. D., Allen, M. D., Bigler, E. D., Goodrich-Hunsaker, N. J., & Hopkins, R. O. (2010). Different patterns of cerebral activation in genuine and malingered cognitive effort during performance on the Word Memory Test. *Brain Injury, 24*(2), 89-99. doi:10.3109/02699050903508218

Lehman, J. & Phelps, S. (2005). West's encyclopedia of American law. Detroit: Thomson/Gale.

Lezak, M.D., Howieson, D.B., & Loring, D.W. (2004). *Neuropsychological Assessment* (4th ed.). New York: Oxford University Press.

Li, H., Rosenthal, R., & Rubin, D. (1996). Reliability of measurement in psychology: From

Loring, D. W., Marino, S. E., Drane, D. L., Parfitt, D., Finney, G. R., & Meador, K. J. (2011). Lorazepam effects on word memory test performance: A randomized, double-blind, placebo-controlled, crossover trial. *The Clinical Neuropsychologist, 25*(5), 799-811. doi:10.1080/13854046.2011.583279

Mahendran, R., Chua, J., Feng, L., Kua, E.H. & Preedy, V.R.. (2015). The Mini-Mental State Examination and Other Neuropsychological Assessment Tools for Detecting Cognitive Decline. *Diet and Nutrition in Dementia and Cognitive Decline*. 1159-1174. 10.1016/B978-0-12-407824-6.00109-9.

malinger (n.d.). In Merriam-Webster.com. Retrieved from http://www.merriam-webster.com/dictionary/malinger.

Mariani, E., Monastero, R., & Mecocci, P. (2007). Mild Cognitive Impairment: A Systematic Review. *Journal of Alzheimer's Disease, 12*(1), 23-35.

McDermott, B. E., Leamon, M. H., Feldman, M. D., & Scott, C. L. (2009). Factitious disorder and malingering. In J. A. Bourgeois, R. E. Hales, J. S. Young, S. C. Yudofsky (Eds.) , *The American Psychiatric Publishing board review guide for psychiatry* (pp. 387-396). Arlington, VA US: American Psychiatric Publishing, Inc.

Merten, T., Bossink, L., & Schmand, B. (2007). On the limits of effort testing: Symptom validity tests and severity of neurocognitive symptoms in nonlitigant patients. *Journal Of Clinical & Experimental Neuropsychology, 29*(3), 308-318. doi:10.1080/13803390600693607

Messé, A., Caplain, S., Pélégrini-Issac, M., Blancho, S., Lévy, R., Aghakhani, N., & ... Lehéricy, S. (2013). Specific and Evolving Resting-State Network Alterations in Post-Concussion Syndrome Following Mild Traumatic Brain Injury. *Plos ONE, 8*(6), 1-10. doi:10.1371/journal.pone.0065470

Miele, A. S., Gunner, J. H., Lynch, J. K., & McCaffrey, R. J. (2012). Are Embedded Validity Indices Equivalent to Free-Standing Symptom Validity Tests?. *Archives Of Clinical Neuropsychology*, *27*(1), 10-22.

Millon, T. (1994). *The Millon Clinical Multiaxial Inventory – III.* Minneapolis, MN: National Computer Systems.

Morel, K. R. (2009). Test Security in Medicolegal Cases: Proposed Guidelines for Attorneys Utilizing Neuropsychology Practice. *Archives of Clinical Neuropsychology*, *24*(7), 635-646. doi:10.1093/arclin/acp062

Morey, L.C. (1991). *Personality Assessment Inventory - Professional Manual.* Florida, USA: Psychological Assessment Resources, Inc.

Novitski, J., Steele, S., Karantzoulis, S., & Randolph, C. (2012). The Repeatable Battery for the Assessment of Neuropsychological Status Effort Scale. *Archives of Clinical Neuropsychology*, *27*(2), 190-195. doi:10.1093/arclin/acr119

Novo, M., Fariña, F., Seijo, D., y Arce, R. (2013). Eficacia del MMPI-A en casos de acoso escolar: Simulación y diagnóstico. *Psychosocial Intervention*, *22*, 33-40. doi:10.5093/in2013a5

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

O'Bryant, SE, Engel, LR, Kleiner, JS, Vasterling, JJ,Black, FW. (2007). Test of Memory Malingering (TOMM) trial 1 as a screening measure for insufficient effort. *Clinical Neuropsychologist*, *21*(3), 511-521. doi:10.1080/13854040600611368

Palmer, I. P. (2003) Malingering, shirking, and self-inflicted injuries in the military. In P. W. Halligan, C. Bass & D. A. Oakley (Eds) *Malingering and illness deception: Clinical and theoretical perspectives* (pp. 52-65). Oxford, England: Oxford University Press.

Pella, R., Hill, B., Shelton, J., Elliott, E., & Gouvier, W. (2012). Evaluation of embedded malingering indices in a non-litigating clinical sample using control, clinical, and derived groups. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists, 27*(1), 45-57. doi:10.1093/arclin/acr090

Powell, M. R., Gfeller, J. D., Hendricks, B. L., & Sharland, M. (2004). Detecting symptom- and test-coached simulators with the Test of Memory Malingering. *Archives of Clinical Neuropsychology, 19*(5), 693-702. doi:10.1016/j.acn.2004.04.001

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in Test-Usage Practices of Clinical Neuropsychologists in the United States and Canada over a 10-Year Period: A Follow-Up Survey of INS and NAN Members. *Archives of Clinical Neuropsychology: The Official Journal of The National Academy Of Neuropsychologists*, *31*(3), 206-230. doi:10.1093/arclin/acw007

Raine, A. (2003) Malingering and criminal behavior as psychopathology. In P. W. Halligan, C. Bass & D. A. Oakley (Eds) *Malingering and illness deception: Clinical and theoretical perspectives* (pp. 52-65). Oxford, England: Oxford University Press.

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status* (*RBANS*). San Antonio, Harcourt, TX: The Psychological Corporation.

Reitan, R. M. & Wolfson, D. (1993). *The Halstead-Reitan Neuropsychological Test Battery: Theory and clinical interpretation* (2nd ed.). Tucson, AZ: Neuropsychology Press.

Resnick, P. J. (1997). Malingering of posttraumatic disorders. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 130–152). New York: Guilford.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Rogers, R. (2001). *Handbook of diagnostic and structured interviewing*. New York: Guilford Press.

Rienstra, Anne & Spaan, Pauline & Schmand, B. (2009). Reference Data for the Word Memory Test. *Archives of Clinical Neuropsychology: The Official Journal of The National Academy Of Neuropsychologists, 31*, 255-62. 10.1093/arclin/acp035.

Rogers, R., Bagby, R.M., & Dickens, S.E. (1992). Structured interview of reported symptoms: Professional manual. Lutz, FL: Psychological Assessment Resources.

Rogers, R., & Bender, S. D. (2003). Evaluation of malingering and deception. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology, Vol. 11* (pp. 109-129). Hoboken, NJ: John Wiley & Sons Inc. doi:10.1002/0471264385.wei1107

Rogers, R., & Neumann, C.S. (2003). Conceptual issues and explanatory models of malingering. In P.W. Halligan, C. Bass, & D.A. Oakley (Eds.), *Malingering and illness deception: Clinical and theoretical perspectives* (pp. 71-82). Oxford, England: Oxford University Press.

Sharland, M. J., & Gfeller, J. D. (2007). A survey of neuropsychologists' beliefs and practices with respect to the assessment of effort. *Archives of Clinical Neuropsychology, 22*(2), 213-223. doi:10.1016/j.acn.2006.12.004

Simon, V., Czobor, P., Bálint, S., Mészáros, A., & Bitter, I. (2009). Prevalence and correlates of adult attention-deficit hyperactivity disorder: meta-analysis. *The British Journal of Psychiatry: The Journal of Mental Science, 194*(3), 204-211. doi:10.1192/bjp.bp.107.048827

Silver, N. C., & Dunlap, W. P. (1987). Averaging correlation coefficients: Should Fisher's z transformation be used? *Journal of Applied Psychology, 72*, 146-148. doi:10.1037/0021-9010.72.1.146

Silverberg, N. D., Wertheimer, J. C., & Fichtenberg, N. L. (2007). [image omitted]An Effort Index for the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). *Clinical Neuropsychologist, 21*(5), 841-854. doi:10.1080/13854040600850958

Slick, D. J., Sherman, E. S., & Iverson, G. L. (1999). Diagnostic criteria for malingered neurocognitive dysfunction: Proposed standards for clinical practice and research. *Clinical Neuropsychologist, 13*(4), 545-561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545

Slick, D. J., Tan, J. E., Sherman, E. S., & Strauss, E. (2011). Malingering and related conditions in pediatric populations. In A. S. Davis (Ed.) , Handbook of pediatric neuropsychology (pp. 457-469). New York, NY US: Springer Publishing Co.

Slick, D. J., Tan, J. E., Strauss, E. H., & Hultsch, D. F. (2004). Detecting malingering: A survey of experts' practices. *Archives of Clinical Neuropsychology, 19*(4), 465-473. doi:10.1016/j.acn.2003.04.001

Smith, G.P. & Burger, G.K. (1997) Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS) *Journal of the Academy of Psychiatry and the Law, 25,* pp. 180–183.

Spearman-Brown to maximal reliability. *Psychological Methods, 1,* 98–107. doi:10.1037/1082-989X.1.1.98

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science, 19*(2), 73-74. doi:10.3758/BF03337426

Stone M. H. & Wright B. D. (1980). *Knox's Cube Test (manual).* Chicago: Stoelting

Strube, M. J. (1988). Averaging correlation coefficients: Influence of Heterogeneity and set size. *Journal of Applied Psychology, 73,* 559-568. doi:10.1037/0021-9010.73.3.559

Sweet, J. J., & Morgan, J. E. (2009). What we currently know about malingering 'to a reasonable degree of neuropsychological certainty' vs. what we would like to know in the future. In J. E. Morgan, J. J. Sweet, J. E. Morgan, J. J. Sweet (Eds.) , *Neuropsychology of malingering casebook* (pp. 557-565). New York, NY, US: Psychology Press.

Tombaugh, T. N. (1996). Test of Memory Malingering. Toronto: Multi-Health Systems.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment, 9*(3), 260-268. doi:10.1037/1040-3590.9.3.260

Toomey, J., Kucharski, L., & Duncan, S. (2009). The utility of the MMPI-2 malingering discriminant function index in the detection of malingering: A study of criminal defendants. *Assessment, 16(*1), 115-121. doi:10.1177/1073191108319713

Ullman, J. B. (2013) Structural Equation Modeling. In B.G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics, (6th Ed)*. Allyn Bacon: New York.

Waddell, G., McCulloch, J., Kummel, E., & Venner, R. (1980). Nonorganic physical signs in low-back pain. Spine, 5(2), 117-125. doi:10.1097/00007632-198003000-00005

Warrington E. K. (1984). *Recognition Memory Test manual*. Windsor, Berkshire: NFER-Nelson

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale Manual*. New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised* (*WAIS-R).* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997a). *Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition* (*WAIS-IV)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2005). *Wechsler Individual Achievement Test 2nd Edition (WIAT II)*. London: The Psychological Corp.

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler Memory Scale (Third Edition).* San Antonio, TX: Psychological Corporation

Wechsler, D. (2009). *Wechsler Memory Scale - Fourth Edition. Manual.* San Antonio, TX: Pearson Assessment.

Wessely, S. (2003) Malingering: Historical perspectives. In P. W. Halligan, C. Bass & D. A. Oakley (Eds) *Malingering and illness deception: Clinical and theoretical perspectives* (pp. 52-65). Oxford, England: Oxford University Press.

Whitney, K. A. (2013). Predicting Test of Memory Malingering and Medical Symptom Validity Test failure within a Veterans Affairs medical center: Use of the Response Bias Scale and the Henry-Heilbronner Index. *Archives of Clinical Neuropsychology, 28*(3), 222-235. doi:10.1093/arclin/act012

World Health Organization.(1948).*International Statistical Classification of Diseases and Related Health Problems, 6th Revision* (*ICD-6*). Geneva: WHO.